University Graduates Tracking Platform: Case Study

Nadir Belhaj [1], Abdelmounaim Hamdane[1], Nour El Houda Chaoui[1] and Moulhime El Bekkali[1]
[1] *Department of Electrical and Computer Engineering of Computing, Sidi Mohamed Ben Abdellah University, Fez, Morocco*

Keywords: Data Warehouse, educational intelligence, graduates performance, graduate tracking, learning analytics, academic analytics, Higher education, data integration, job market entry.

Abstract: In this paper, we will explain our approach in building a graduates tracking platform, which will enable a detailed analysis of the university graduates, their labour market entry, hiring companies, industries, and sectors. We implemented a data warehouse to track university graduates and analyze their career paths after graduation. This analysis is used for university courses assessment and measuring the demand level of skills we teach our students in the labour market.

1 INTRODUCTION

Tracking graduates, assessing the education system and gathering feedback about courses and university student's life is becoming a significant need for every educational institution to make intelligent and strategic decisions based on an extensive volume of data collected every year (Moscoso-Zea et al., 2016). In need of being able to enhance the quality of courses and better serve the students with the right skills needed in the labour market, we implemented a data warehouse inside of our university Sidi Mohammed Ben Abdellah, located in Fez city (Wierschem et al., 2003), (Bouaziz et al., 2017). In Morocco, none of the universities implemented a complete students and graduates data warehouse capable of storing data, analyzing and delivering fast and accurate reports. A data warehouse is a collection of data from various sources stored in a large database then processed into a multi-dimensional storage form to make it easy for querying and reporting (Sulianta and Juju, 2010), (Gosain and Heena, 2015) and (Moscoso-Zea et al., 2016). The Labour market is highly dynamic because of competition, growth, and demand of customers; this is why the university nowadays needs to prepare the students for a never-stable environment with the correct skills, techniques, training, and tools. To be capable of doing this, establishing the right strategies and processes based on the right data is needed. Having a data warehouse full of information's about our student's and their after university career data helps in better understanding the university students culture by performing data mining and data analysis to learn more about the students and get answers for questions such as: where they prefer to live after graduation, are they willing to relocate or not, how much does it take to land a job after graduation, how many interviews needed by job vacancy, what is the median salary by sector, are the skills learned in the university in demand in the labour market and many more, (Moscoso-Zea et al., 2016; Buenstorf et al., 2016; Bichsel, 2012).

This research was held in Sidi Mohamed Ben Abdellah University in Morocco with the collaboration of faculty, institutes, students, and graduates and focuses on data warehouse design and data collection over the span of four years.

[1] https://orcid.org/0000-0001-9179-0295
[2] https://orcid.org/0000-0001-7645-1287
[3] https://orcid.org/0000-0002-4228-035X
[4] https://orcid.org/0000-0002-1098-6841

Belhaj, N., Hamdane, A., Chaoui, N. and El Bekkali, M.
University Graduates Tracking Platform: Case Study.
DOI: 10.5220/0010736300003101
In *Proceedings of the 2nd International Conference on Big Data, Modelling and Machine Learning (BML 2021)*, pages 451-456
ISBN: 978-989-758-559-3

2 METHODS

2.1 The Student Career Tracking Data Warehouse Architecture

Building a data warehouse has always been an important decision for every enterprise and organization, including universities. The most critical decision in its design is finding the right way to follow to build it, whether a Bill Inmon Top-down approach which advocates that a global data warehouse is constructed first and serves as a basis for small data marts (Inmon, 2005). Or to follow a Bottom up data warehouse design approach recommended by Ralph Kimball (Kimball and Ross, 2013), based on building data marts first to provide the reporting and analytics capability for specific business processes and then compounding them to make a data warehouse also named dimensional modelling (Kimball and Ross, 2016).

The university is composed of faculty, institutions and departments that operate separately and independently. This is why a bottom-up Kimball's approach is recommended for implementing our data warehouse (Vogelgesang and Appelrath, 2016).

We will begin first by gathering the primary data sources that will connect to an ETL (Extract Transform Load) tool to clean, unify and load data in data marts that answers some specific questions and through a bus architecture we will chain them up then build the data warehouse.

In Figure 1 below, we explain the process of extracting data from our primary data sources, which are the university database that is connected to the data collection platform the university web application that collects data from students and graduates. StudentDB a database that contains all students path before university data such as personal data, high school data, family data, etc, and a bunch of files containing data about the university, different study programs, which are subject to change and delivered every year from the Moroccan minister of research and higher education.

Through an ETL process, we clean and unify our data to load it in the data marts and extract reports.

We decided to gather the most important questions that we need to understand the pattern of our graduates then collected a bunch of inquiries like the following:

Q 1: How many graduates by faculty and by degree or specialty?

Q 2: What is the average mark of our graduates by the institute and by degree?

Q 3: How many graduates are hired within the first six months after graduation?

Q 4: Which are the most hiring sectors of our graduates?

Q 5: What is the average time for our graduates to find a Job?

Q 6: How many graduates used to work while studying?

Based on these questions, we started designing our data marts, extracted three principal data marts based on three main events: enrollment, graduation, and hiring (Rahman et al., 2015):

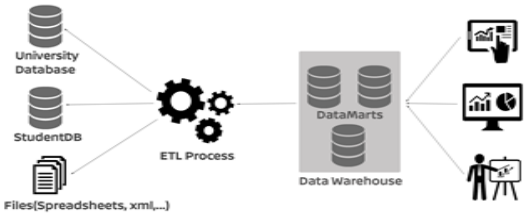

Figure 1. Data extraction, transformation, and loading in Bottom-up approach

2.2 Design of Data Marts

2.2.1 Defining the Scope of Data Marts

Enrolment Data Mart

The first data mart is the enrollment data mart, which provides access to meaningful data that is specific to the student registration phase.

This data mart will provide answers to some specific questions such as:

- How many students enrolled in institute X and degree YY?
- Where our students are coming from?
- How many foreign students per institute and degree?
- What is the popularity of each degree?
- How many foreign students overall?

Logical design of enrollment data mart

A logical design is a conceptual design, which is highly abstracted from the physical layer, and it is called dimensional modelling in data mart design and Ralph Kimball first introduced this concept in (Kimball and Ross, 2013). We begin by defining a central fact table that models an event which can be a single transaction such as enrollment by a student, a periodic time where a snapshot of events are collected such as registered students in spring session or a

business process with a clear beginning and end. Every fact table has a set of associated dimensions tables that contain information of the fact table in the form of entities (students, institutes, courses, etc.).

Classifying Data for the data mart schema

To classify our extracted data in facts and dimensions tables, we need an appropriate schema or data model and we chose the star schema as it offers fast querying, load performance and ease of understandability and navigation (George et al., 2015). One standard methodology in star schema's design is to start with the creation of dimensions first, then the fact table and we arranged our dimensions for the enrollment data mart as follows:

- Student Dimension
- Institute Dimension
- Degree Dimension
- Date Dimension

This dimension contains data about our university-enrolled students. Every row is related to a unique student with attributes providing information about its personal data (Full Name, Gender, Email, etc.), previous work and studies, family conditions and other details. This dimension is helpful in performing analyses of newly enrolled students and graduates achievements related to their past work, studies and their living conditions.

Institute Dimension:

The Institute dimension embodies every faculty and institute attached to Sidi Mohammed Ben Abdellah University, which is composed as of this date of X Institute, YY Faculty, ZZ. Every row contains Institute Name, Type, address, and other details that can help in the process of filtering our reports by institute, faculty and other attributes.

Degree Dimension:

Every student is enrolled in a particular study program and this dimension gathers every detail about each degree in our university such as title, level (PhD, Masters, etc…) , type (scientific, literature, etc.) , duration and other details. Each attribute will be helpful in extracting precise information about student's study enrollments in every study program.

Date Dimension:

We need to query our report annually, quarterly, monthly, weekly and daily and here resides the importance of our date dimension.

This dimension is standard to all data marts and will be drawn on the enrollment data mart alone till we chain all of them to form the entire data warehouse.

Enrollment Fact Table:

The fact table contains the metrics of our data mart, all the fields that we want to summarize and foreign keys of our dimensions. Enrollment fact table has two outlined attributes, which are the high school graduation mark and the paid fee for professional degrees.

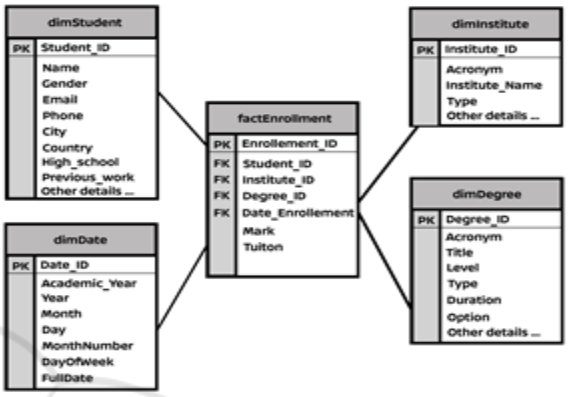

Figure 2: Star schema of Enrollment data mart with the fact enrollment table and its dimensions

Graduate Data Mart

The scope of graduate's data mart is answering some of the essential questions that are related to the graduation of every student and can provide meaningful insights about their internships, future plans, endeavours, and accurate feedback about life inside the university and residency. This data can help to understand the correlations between graduates performance in the job market and their progress through their study cursus.

Example of the graduate's data mart queries:

Q1: How many graduates are satisfied by the university training, dormitory, etc ?

Q2: How many students used to follow a paid training while studying outside of the university?

Q3: What is the average of students that wants to follow their studies inside the same university and abroad?

Q4: How many graduates intend to start their job (launch a start-up)?

Q5: how many scientific, literature, finance, etc students?

Q6: how many graduates used to work (freelancing) while studying?

Q7: what is the average of English speaking graduates?

Q8: How many graduates with more than a diploma?

Q8: How many graduates know how to use a computer and essential applications?

Q9: How many graduates do have a smartphone; use a smartphone instead of a computer in their studies?

2.2.2 Dimension Tables

The graduate's data mart star schema is consisting of the following dimensions:

- -Student Dimension (same table used students data mart)
- Internship Dimension
- Date Dimension (same table used in students data mart)
- Institute Dimension (same table used in students data mart)
- Degree Dimension (same table used in students data mart)
- Futur Plans Dimension
- Feedback Dimension

Internship Dimension:
This dimension describes every internships for each graduate and can help in understanding the relationship between the graduate's future job and his past internships. Each row contains company name, project title, duration, company sector, country and other details.

Future Plans Dimension:
The future plans dimension describes the university graduates goals and endeavors. Simplifies the comparison of the actual state of every graduate and his pre-graduation vision and helps understanding the correlations between his job and his plans. It embodies attributes like First-year goals, second-year goals, five-year goals, if he is willing to launch his own business, etc.

Feedback Dimension:
University Sidi Mohammed Ben Abdellah is among top universities in Morocco and top 500 in the world and to achieve more it prioritizes a quality first rule. That is why we need more feedback from all the university stakeholders.

Graduation Fact Table:
The graduation fact table represents the event of graduation of each student in the university and holds foreign keys to the mentioned dimensions, date of graduation and one measurable attribute, which is the grade or mark.

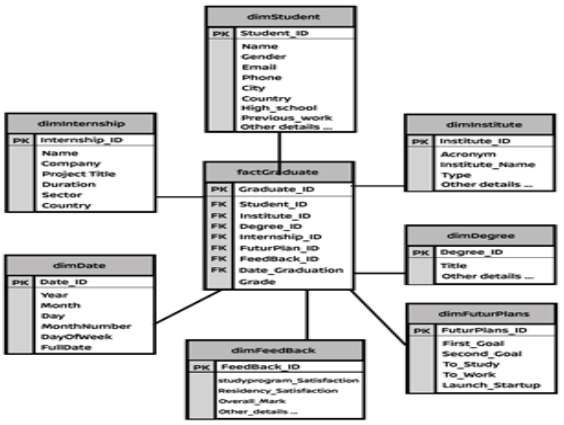

Figure 3: Star schema of Graduates data mart with the fact graduate table and its dimensions

Job Data Mart

This data mart is the missing component in our career tracking data warehouse. Based on our collected data, it serves as a primary source for analyzing the labour trends, university graduates performance, graduates skills compatibility with Job market demand and a lot more.

Analyses through this data mart will open space for more insights and provides answers to many questions such as:

Q1: What are the top hiring companies of our graduates?

Q2: What are the hiring industries of our graduates?

Dimension Tables
The Job Data Mart is composed of four dimensions:

- Student Dimension (same table used in graduate and enrollment data mart
- Employer Dimension
- Sector Dimension
- Contract Dimension
- Date Dimension (same table used in graduate and enrollment data mart).

Employer Dimension:
The employer dimension provides information about each employer such as the company name, size type, sector, desired skills, and profiles, advantages, etc. Each row represents a unique employer.

Sector Dimension:
This dimension contains information about the diverse sectors and industries available in today's labour such as industry, photography, art, film making, Information systems, administration, etc. Available attributes help in the analysis of current market trends and repartition of our graduates in the different country sectors and international labour market, every sector is divided into sub-sectors, which will provide a drill-down and drill up inside our reports.

Contract Dimension:
In Morocco, and many other countries we have a different type of job contracts (Full Time, Part Time, permanent contract, temporary contract) and having this detail as a dimension will enable querying data and analyses based on the type of contracts.

Job Fact Table:
The Job fact table models the transition from a graduate to a hired person event and holds foreign keys to the above-seen dimensions, date of hiring and one metric attribute, which is the salary.

3 RESULTS

The student career development data warehouse can help us get meaningful insights by a simple drill across reports. We can as example get the number of applicants, how many of them got accepted and enrolled, average students mark, how many did graduate and the most critical numerical value that our data warehouse provides is how many of our students got hired.

Table 1: simple drill-across report

Institute ZZZ

Academic Year	Applicants	Enrolled	Average Mark	Graduated	Hired
2014/15	500	380	14,25	318	280
2015/16	540	418	14,64	379	310
2016/17	610	460	14,66	436	390
2017/18	812	519	15,01	480	353

Table 2: Report refinement by drilling down

Institute ZZZ

Academic Year	Degree	Applicants	Enrolled	Average Mark	Graduated	Hired
2014/15	Degree X	200	100	14,20	80	65
	Degree Y	300	150	14,02	130	96
2015/16	Degree X	230	120	14,23	100	80
	Degree Y	310	170	14,08	150	114
2016/17	Degree X	250	130	14,50	115	90
	Degree Y	320	180	14,30	165	120
2017/18	Degree X	260	140	14,60	127	94
	Degree Y	340	190	14,35	175	130

4 CONCLUSIONS

We explained the design of a student career progression data warehouse that can be implemented following bottom up approach by defining the basics data marts and connecting them through common dimensions to from a bus architecture. This architecture is capable of providing a solid data warehouse that can be queried to get important information's about how our graduates progressed in the labour market and how the study program of the university helped them in reaching their career goals. This study and work is a basis for new research in Big data, data mining and machine learning by adding more data sources, exploring new data patterns, and extracting better reports and analysis.

REFERENCES

Oswaldo Moscoso-Zea, Andres-Sampedro, Sergio Luján-Mora, 2016, *"Datawarehouse design for educational data mining," 15th International Conference on Information Technology Based Higher Education and Training (ITHET) 8-10 Sept.*

David Wierschem, Jeremy McMillen, Randy McBroom, 2003, *"What Academia can gain from building a Data Warehouse", Number 1, EDUCAUSE QUARTERLY.*

S. Bouaziz, A. Nabli, and F. Gargouri, 2017, "*From Traditional Data Warehouse to Real Time Data Warehouse," in International Conference on Intelligent Systems Design and Applications.*

F. Sulianta and D. Juju, 2010, *"Data Mining. Jakarta: PT. Elex Media Komputindo.*

Gosain and Heena, 2015, *"Literature Review of Data model Quality metrics of Data Warehouse," in International*

Conference on Intelligent Computing, Communication & Convergence, pp. 236–243.

O. Moscoso-Zea, Andres-Sampedro and S. Luján-Mora, 2016, *"Datawarehouse design for educational data mining," 15th International Conference on Information Technology Based Higher Education and Training (ITHET), Istanbul, pp. 1-6. doi: 10.1109/ITHET.2016.7760754.*

Guido Buenstorf, Matthias Geissler, Stefan Krabel, 2016, *"Locations of labour market entry by German university graduates: is (regional) beauty in the eye of the beholder", February, Volume 36, Issue 1, pp 29–49 Springer Berlin Heidelberg. https://doi.org/10.1007/s10037-015-0102-z*.

W. H. Inmon, 2005, *"Building the data warehouse. New York: John Wiley & Sons"*.

R. Kimball and M. Ross, 2016, *"Fact Table Core Concepts," in the Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence.*

T. Vogelgesang and H.-J. Appelrath, "2016, *PMCube: A DataWarehouse-Based Approach for Multidimensional Process Mining," in International Conference on Business Process Management, pp. 167–178.*

Ralph Kimball and Margy Ross, 2013, "The Data Warehouse Toolkit, 3rd Edition" *Wiley.*

J. George, B. V. Kumar, and S. Kumar, 2015, "*Data Warehouse Design Considerations for a Healthcare Business Intelligence System," Proc. World Congr. Eng., vol. 1.*

L. Rahman, S. Riyadi, and P. Eko, "2015, "Development of Student Data Mart using Normalized Data Store Architecture," in Advanced Science Letters, pp. 3226–3230.

J. Bichsel, 2012, "Analytics in Higher Education Benefits, Barriers, Progress and Recommendations," [Online]. Available: https://net.educause.edu/ir/library/pdf/ERS1207/ers1207.pdf. [Accessed: 26-Mars 2021].