

Decision-Making Information System for Academic Careers in Congolese Universities: From Analysis to Design of a Data Warehouse

Boribo Kikunda Philippe^{1,2,3}, Thierry Nsabimana¹, Longin Ndayisaba¹, Jules Raymond Kala², Jérémie Ndikumagenge¹, Elie Zihindula Mushengezi³

¹Centre de Recherche en Infrastructure Environnement et Technologie (CRIET), Faculty of Engineering Sciences, Université du Burundi, Bujumbura, Burundi

²Faculty of Sciences, Université Catholique de Bukavu (UCB), Bukavu, DR Congo

³Management Computer Département, Institut Supérieur Pédagogique de Bukavu (ISP/Bukavu), Bukavu, DR Congo

Email: boribokikunda@gmail.com, kikunda.boribo@ucbukavu.ac.cd, thierry.nsabimana@ub.edu.bi,

longin.ndayisaba@ub.edu.bi, raymondkala1@gmail.com, jeremie.ndikumagenge@ub.edu.bi, eliezihindula@yahoo.fr

How to cite this paper: Philippe, B.K., Nsabimana, T., Ndayisaba, L., Kala, J.R., Ndikumagenge, J. and Mushengezi, E.Z. (2023) Decision-Making Information System for Academic Careers in Congolese Universities: From Analysis to Design of a Data Warehouse. *Open Journal of Applied Sciences*, 13, 2395-2407.

<https://doi.org/10.4236/ojapps.2023.1312187>

Received: November 13, 2023

Accepted: December 23, 2023

Published: December 26, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Universities collect and generate a considerable amount of data on students throughout their academic career. Currently in South Kivu, most universities have an information system in the form of a database made up of several disparate files. This makes it difficult to use this data efficiently and profitably. The aim of this study is to develop this transactional database-based information system into a data warehouse-oriented system. This tool will be able to collect, organize and archive data on the student's career path, year after year, and transform it for analysis purposes. In the age of Big Data, a number of artificial intelligence techniques have been developed, making it possible to extract useful information from large databases. This extracted information is of paramount importance in decision-making. By way of example, the information extracted by these techniques can be used to predict which stream a student should choose when applying to university. In order to develop our contribution, we analyzed the IT information systems used in the various universities and applied the bottom-up method to design our data warehouse model. We used the relational model to design the data warehouse.

Keywords

Data Warehouse, University Courses, Universities of South Kivu

1. Introduction

A wide range of student data circulates within the university. They are handled

by the academic services office, the jury office, the faculty office, the alumni office and so on. As a result, several data bases coexist within the university. What's more, these data are stored in a variety of sources such as paper, Excel sheets, Access or MySQL files. They are used separately and operationally by each of these departments.

The fact that these data are not used effectively as a single whole means that the reality contained within them cannot be understood. This makes it difficult to explain certain phenomena, such as success, drop-out or failure in certain options. For example, computer science is one of the most popular fields of study at local universities. Unfortunately, there is a surprising phenomenon. After the first year, eighty percent of students hate the basic computer science course, in this case programming. We don't know why.

That's why these students need to be able to make informed decisions about their choice of course of study, which will also determine their choice of career later on. It is essential to build a decision support system for this purpose. Unlike a transactional information system, which can only process operational data in real time, a decision-support system focuses on the analysis and presentation of historical and integrated data to support strategic decision-making.

Google, for example, uses a wide range of data, including user behavior, search history and content analysis, to deliver the most relevant search results to users. This has helped it to become the most popular search engine and generate considerable revenue. Another example is Amazon. It has a system that uses data to understand customer behavior and preferences, enabling it to offer personalized recommendations and targeted advertising. This helps it to increase sales and improve customer satisfaction. Universities can also improve their offer to students by extracting knowledge from historical data on student career paths. This could lead to a reduction in failures, for example.

Consequently, we need a means of collecting, organizing and exploiting this data. The aim of this article is therefore to design and implement a data warehouse for students' academic careers. This tool will make it possible to collect, centralize and archive data on students' academic careers from the six major universities in South Kivu Province, and to transform them for analytical purposes. Data mining is not currently involved. This latter subject will be part of future work.

Rather, the aim is to make available a database of student profiles and their academic history, from admission to university through to graduation or withdrawal. To do this, we will go through two phases: 1) the design of the data warehouse and 2) the restitution or exploitation phase. For the first phase, we will propose the sub-phases of data warehousing, data feeding and dimensional modeling based on the fact/dimension method. As for the restitution phase, we will use appropriate tools such as SQL server to make the best use of the data already stored in the warehouse.

Apart from the introduction, in the first section we will present the theory behind data warehouse technology. We will present the properties, construction

phases and types of data warehouse modeling. The second section presents the methodology used. The third section will focus on analyzing existing data sources in order to identify all the facts to be analyzed. Finally, we will present the results obtained before drawing a conclusion.

2. Theory on Data Warehouses

2.1. Properties of a Data Warehouse

The classic definition of a data warehouse is that of [1]:

“The Data Warehouse is a collection of subject-oriented, integrated, non-volatile, historical data organized to support a decision-making process.

We can describe these properties in detail [2]:

- Subject-oriented: warehouse data is organized by fact or subject rather than by application.
- Historicized: the evolution of data is essential for decision-making, for example, by relying on past evolutions to predict future ones. This is a layered model.
- Non-volatile: data inserted into a warehouse is never modified or deleted, *i.e.* it is permanent and cannot be modified, and is retained for future analysis.
- Integrated: data comes from different sources before being stored in the warehouse. Integration is important to ensure that formats, for example, match and that information is consistent.

In contrast, the data properties of an operational or transactional database are: application-oriented, volatile, with an organization system that enables queries to be processed rapidly. This shows that a data warehouse is a decision support tool for a decision-maker, whereas a transactional database is adapted to the business level.

2.2. Warehouse Construction Phases

The construction of a data warehouse is subdivided into 3 main stages [3]:

- Analysis of data sources
- Design of the warehouse data model
- Development of analysis tools

Each phase has its own set of issues and tools. In the first phase, good tools are needed to load and integrate data from several sources. To achieve this, either tools commonly known as Extract Transform Load are used, or, thanks to a procedural language such as C and a declarative language such as SQL, these operations can be programmed manually. Most ETL tools are quick and easy to use, but they are not free. They are best suited to non-IT end-users. For data engineers, they can easily program data extraction, transformation and loading operations manually [4].

In the second phase, the main challenge is data organization. It is crucial to have a good representation of the data to facilitate the materialization of views, good data management and warehouse refreshment. Several models exist: the

star model, the vial model, the constellation model, etc. [1].

As for the last phase, analysis or interrogation, good tools such as OLAP¹ can help you achieve this goal [3]. Of course, most data warehouses designed for use by decision-makers require integrated OLAP-based analysis tools. We find this in intelligent business applications [2].

A data warehouse can be exploited by different technologies, such as OLAP, Reporting and Data mining. It's important to note that, for the purposes of this study, our focus is on data mining. From this point of view, the warehouse serves only as a data collection tool, which will later be exploited by artificial intelligence or machine learning algorithms.

Today, there are many platforms available for managing data warehouses. Thanks to the Pentaho ecosystem, building a data warehouse is much easier. Pentaho Data Integration is the open-source tool for creating data warehouses with data loading and integration procedures.

2.3. Data Warehouse Modeling

The data structure of a transactional application is based on a relational model. This formalism is well suited to transactions, *i.e.* CRUD² operations.

On the other hand, to design a data warehouse, we use a dimensional model [5]. This comprises a fact table linked to dimension tables by a join. This is a star schema. So, a data warehouse is a set of fact tables.

There are three types of modelling [6]:

- Star pattern: In this type of model, we find a central fact linked to dimensions.
- Snowflake: This model derives from the previous one, with a central fact and around it the various dimensions are broken down into sub-hierarchies.
- Constellation: This is a model containing several facts that share common dimensions.

These three types of modeling are based on relational formalism, *i.e.* the use of foreign key joins. However, there is another type of modeling called multidimensional (cube) modeling, which consists in representing the subject to be analyzed as a point in a multidimensional space, which we won't be using in this study. The axes of projection are the dimensions, and the value of the measurement is the point in space.

3. Methodology

In order to design and implement a first prototype of our data warehouse, the methodology developed in this article is that used by Ralph Kimball, who advocates designing dimensional structures without having to create the warehouse entirely [7]. The result of this process can then be used in subsequent steps as the basis for modeling dimensional extracts [3]. This is a faster and simpler way of

¹Online Analytical Processing allows users to analyze data from multiple databases at the same time.

²Create Read Update Delete are common operations of a transactional or operational database.

creating a data warehouse. This basic “bottom-up” approach guides the whole project methodology.

Kimball’s bottom-up approach is generally based on the analysis of user needs. But it is usually coupled with another goal-driven analysis, to make it sustainable in the face of organizational change and dynamism [8]. It’s a hybrid analysis. From this analysis stems the identification of the facts to be analyzed and the dimensions associated with them. At this level, several types of model are possible. Our choice will be a fact/dimension model without a pre-processing zone, as the data will not undergo any major modification when loaded.

Finally, it will be important to specify the warehouse model to be used for its physical storage [9]. In contrast to the cube model, we’re going to choose the relational model, ROLAP, which will enable us to respond more easily to the need for *data mining in the future*. To this end, we’re going to use sqlserver 8 to implement the data warehouse.

4. Analysis of Existing Situation

Analysis is a crucial step in the design or representation of an object. In this section, we will analyze existing data sources actually used in universities. This will enable us to identify the properties of the different data sources in relation to our chosen approach.

4.1. Data Sources

In this section, we describe the information systems used in the various universities in relation to services. For us, these are a source of data for our data warehouse. More specifically, these are the schooling department, the jury and deliberation department, and the data archiving and report card and diploma delivery department.

4.2. Description of Sources

GP7 file (Management of Programs, Courses and Students) Developed by the University of Lubumbashi in collaboration with the University of Mons as part of the CUD’s UniversiTIC program and the VLIR-UOS, this file is used to manage programs, courses and students. A computerized academic management application that manages registration, curriculum and the deliberation process [10].

File characteristics:

- Owner: University of Lumbubashi
- File type: C #
- Year: 2004
- Method of production: UniversiTIC project
- Updated: Annual
- Size: 1 GB

Académia file. This is a web application developed by ISP’s IT department since 2015 to register students.

- Owner: ISP
- File type: Php-Mysql
- Year: 2015
- Updated: Annual
- Size: 50 MB

GS-Scholar (Scholar Management System) Web application deployed within the institution to manage student enrollments, control schooling, manage admissions to classes (initial and rising), and produce temporary and definitive student lists.

- Owner: ISDR
- File type: Php-Mysql
- Year: 2021
- Updated: Annual
- Size: 60 MB

NoteLMD

Two sources are external: data from the EPSP (provincial ministry...) and data from the alumni system.

Fichier Alumni is a web application that collects information on former students in relation to what they studied at university and what they became in real life after their studies. This application is only implemented at the Catholic University of Bukavu. This may be useful for studying the profiles of students who succeed in integrating into professional life in relation to specialization, background, etc., for example.

Fichier EPST is an Excel file that stores information on all the province's schools and their characteristics, as well as information on all primary and secondary school finalists. This data is regularly centralized in EPST software in the capital.

We note that there are several data sources that manipulate student data at different universities as we see in **Table 1**. A data warehouse model is therefore appropriate for collecting, organizing and historicizing these data.

5. Results and Discussion

We are currently studying how to design a tool that will enable us to collect data on the academic background of students at universities, particularly those in South Kivu. We hypothesized that the data warehouse approach would be well suited to solving this problem, compared with the transactional database approach. To do this, we illustrate in **Table 2** the advantages of data warehouses compared to databases.

Setting up a data warehouse involves three key elements: ETL, data warehouse structure and operating tools. We will present the three components in relation to our study.

5.1. ETL

Once the warehouse has been created, the Pentaho Data Integration tool from

Table 1. File used in relation to the institution.

	School Department	Jury and deliberation department	Archiving and delivery of newsletters and diplomas
Université Catholique de Bukavu (UCB)	GP7	GP7	- Grids of deliberations archived in paper - Word file for gradebook delivery
Institut Supérieur Pédagogique de Bukavu (ISP)	Academia	Academia	- Grids of deliberations archived in paper - Word file for gradebook delivery
Institut Supérieur en Développement Rural (ISDR)	SGS-Scholar	NoteLMD	- NoteLMD - MS Excel workbook with VBA scripts
Institut Supérieur des Techniques Médicales (ISTEM)	MS Excel workbook with VBA scripts.	MS Excel workbook with VBA scripts.	- Grids of deliberations archived in paper - Word file for gradebook delivery
Université Officielle de Bukavu (UOB)	MS Excel workbook with VBA scripts.	MS Excel workbook with VBA scripts.	- Grids of deliberations archived in paper - Word file for gradebook delivery
Université Evangélique en Afrique (UEA)	MS Excel workbook with VBA scripts.	MS Excel workbook with VBA scripts.	- Grids of deliberations archived in paper - Word file for gradebook delivery

Table 2. Comparative advantages of the warehouse versus existing information systems.

IS ³ Existing academics	IS Data warehouses
Contains raw data that will require complex queries to arrive at decisions. We will speak of “information”.	Contains information useful for decision-making. We refer to this as “knowledge”.
Closed systems, operating independently.	Centralization of data from the organization’s various systems.
Relatively average data size.	Large volume of data.
Relational structure based on the entity-association model.	Dimensional structure based on the fact-dimension model.
Unintuitive understanding of data	Intuitive understanding of data.
Highly normalized tables	Non-standardized tables
High security of SGBG, which causes slow access to system data.	Once loaded into the system, access is quick.
Add, delete, modify operations	Only the add operation.
Only detailed and up-to-date Data	Includes detailed, up-to-date and historical data, summary data (aggregated, summarized).
Many tables, join which implies a long search time.	Fewer tables and joins, which implies an optimization of response time when using data.

the Pentaho platform is used to load data into it. It enables data manipulation and transformation operations to be designed and executed. Thanks to a step-based graphical model, it is possible to create processes consisting of data imports and exports, and various transformation operations such as conversions, joins and the application of filters, without programming.

Indeed, Pentaho Data Integration lets you do exactly what the most skilled

³Information system.

code developers can accomplish, in a fraction of the time and without the need for manual coding. It supports a wide range of data sources, including over 30 open source and proprietary database platforms, as well as flat files.

Other tools that can be evaluated for this task include JasperSuite from JasperSoft, SAS, Data Manager (Cognos), Cognos8 (Cognos), PowerCenter (Informatica) and Crystal Decisions from Business Objects (BO). Unfortunately, most of them are still not free. A comparative study of the different tools may be essential to guide the choice of one over the other.

5.2. Warehouse Structure

Secondly, it is important to present the structure of the warehouse obtained. This was obtained using the bottom-up approach. With this warehouse, two major objectives can be achieved, since it possesses data.

- To help students choose a course of study that best suits their academic background, skills and interests, and finally their career goals. The warehouse will therefore store past (background), present (grades obtained) and future (career goals) data. It could thus play a key role in developing a system for recommending an option to students.
- Help institutions to better understand the factors that influence student success in different courses of study.

The structure of our warehouse is shown in **Figure 1**, taking into account the purposes described above. The dimension tables describe the characteristics of students, streams, courses, teachers and universities. Foreign keys are used to link the fact table to the dimension tables. The fact table contains measures that will play the role of a variable to be modeled in order to produce predictive models.

Also, we can point out that by using the university dimension in this model, information characterizing the universities of South Kivu will also be stored. This will make it possible to analyze students' choices of these universities on the basis of their financial means. It should be noted that most universities in South Kivu operate on the basis of what students pay.

5.3. Exploitation

Unlike existing university information systems, a warehouse offers the following advantages [1]:

- Let's assume that T_i is the access time to an item of information in the existing University IS and T_e is the access time to an item of information in a data warehouse system;
- Let T_1 also be the time it takes to find information in the existing university IS, due to the fact that the system is made up of several independent transactional systems;
- Let T_2 be the time associated with the existence of several tables and joins in transactional systems;

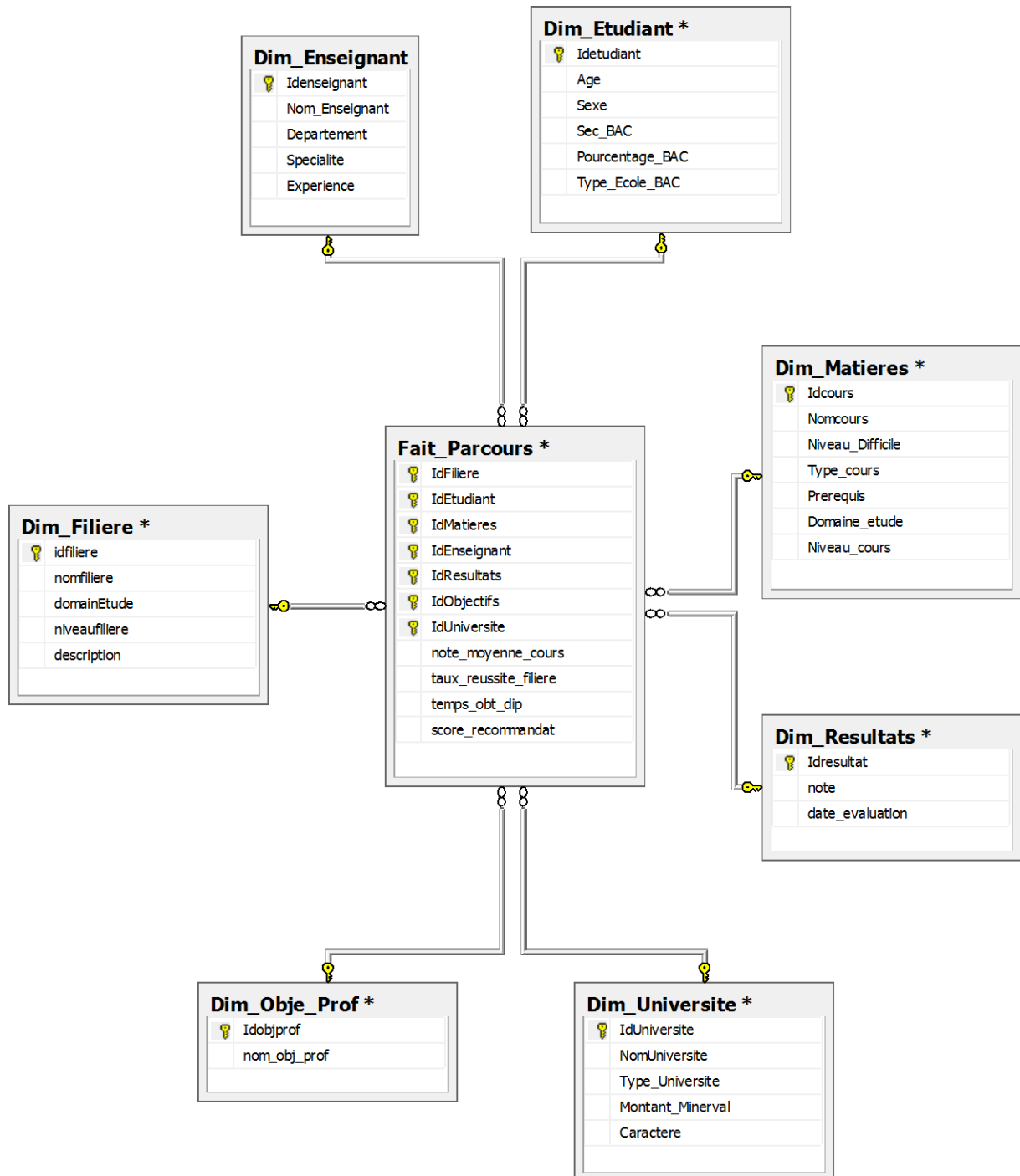


Figure 1. Logical data model.

- Let T_3 denote the time required to verify access rights to a piece of information during a search in transactional systems;
- The total time to access information in the existing (operational) university system is given by the following formula:

$$T_t = T_1 + T_2 + T_3$$

We know that $T_e < T_2$ because T_e contains fewer tables and fewer joins.

Clearly, the time it takes to access information in a warehouse is much shorter than in a transactional system. What's more, data is historized. The data organization model for analysis purposes is therefore that shown in **Figure 1**.

Data ready for analysis will be obtained using an sql query of the form:

```
SELECT * FROM Fait_Parcours INNER JOIN Dim_Etudiant ON Fait_Parcours.IdEtudiant=Dim_Etudiant.IdEtudiant INNER JOIN Dim_Filiere ON Fait_Parcours.IdFiliere=Dim_Filiere.IdFiliere INNER JOIN Dim_Universite ON Fait_Parcours.IdUniversite=Dim_Universite.IdUniversite;
```

And which gives the results in the form of a data set illustrated in **Figure 2**.

The columns represent the characteristics and the rows represent the students. We show how we can obtain other data sets which can also be the subject of an analysis in **Table 3**. The first column gives the description of the query and the second column specifies it in SQL.

Table 3. Some query in SQL.

Query	Query in SQL
A query allowing you to select all the students of a university with their information.	SELECT * FROM Fait_Parcours inner_join Dim_Etudiant ON Fait_Parcours.IdEtudiant=Dim_Etudiant.IdEtudiant INNER JOIN Dim_Filiere ON Fait_Parcours.IdFiliere=Dim_Filiere.IdFiliere INNER JOIN Dim_Université ON Fait_Parcours.IdUniversité =Dim_IdUniversité WHERE IdUniversité=1;
A query that displays the profile of all students in a given course.	SELECT * FROM Fait_Parcours inner_join Dim_Etudiant ON Fait_Parcours.IdEtudiant=Dim_Etudiant.IdEtudiant INNER JOIN Dim_Filiere ON Fait_Parcours.IdFiliere=Dim_Filiere.IdFiliere WHERE nomfiliere = 'AGRONOMIE'
A query that allows you to display the profile of students with an average score of less than 50 percent in a given field	SELECT * FROM Fait_Parcours inner_join Dim_Etudiant ON Fait_Parcours.IdEtudiant=Dim_Etudiant.IdEtudiant INNER JOIN Dim_Filiere ON Fait_Parcours.IdFiliere=Dim_Filiere.IdFiliere WHERE nomfiliere = 'AGRONOMIE' And note_moyenne<50
A query that allows you to have students who have achieved a score in the past and who are performing well in a given sector	SELECT * FROM Fait_Parcours inner_join Dim_Etudiant ON Fait_Parcours.IdEtudiant=Dim_Etudiant.IdEtudiant INNER JOIN Dim_Filiere ON Fait_Parcours.IdFiliere=Dim_Filiere.IdFiliere WHERE nomfiliere = 'MEDECINE' And sec_Bac='SCIENTIFIQUE' AND pour_BAC>=60
A request allowing you to have the profile of students who had a goal and who chose a given sector	SELECT * FROM Fait_Parcours inner_join Dim_Etudiant ON Fait_Parcours.IdEtudiant=Dim_Etudiant.IdEtudiant INNER JOIN Dim_Filiere ON Fait_Parcours.IdFiliere=Dim_Filiere.IdFiliere INNER JOIN Dim_Obj_Prof ON Fait_Parcours.IdObjprof=Dim_Obj_Prof.Idobjprof WHERE nomfiliere = 'MEDECINE' And sec_Bac='SCIENTIFIQUE' AND pour_BAC>=60

ant	Age	Sexe	Sec_BAC	Pour...	Type_E...	idfilere	nomfilere	domainEtude	description	Id...	NomUniversite	Type_Universite	Mont...	Caractere
1	18	M	Bio-Chimie	62	collège	01	Droit	Sc Sociales	études en sciences juridiques	01	UCB	Université	650	privée
2	19	F	Commercial	65	Lycée	01	Droit	Sc Sociales	études en sciences juridiques	01	UCB	Université	650	privée
3	18	M	lettres	69	collège	01	Droit	Sc Sociales	études en sciences juridiques	02	UEA	Université	500	privée
4	26	F	lettres	50	collège	02	Informatiques	Sc Technologies	études en informatique	03	ISP	Institut	300	public
5	21	M	Commercial	75	complexe	02	Informatiques	Sc Technologies	études en informatique	02	UEA	Université	500	privée
6	19	M	lettres	70	Lycée	02	Informatiques	Sc Technologies	études en informatique	03	ISP	Institut	300	public
7	18	F	Math-Phys	63	collège	03	Médecine	médecine	études en médecine	04	UOB	Université	400	public
8	17	F	Commercial	80	institut	03	Médecine	médecine	études en médecine	04	UOB	Université	400	public

Figure 2. Data set for data analysis.

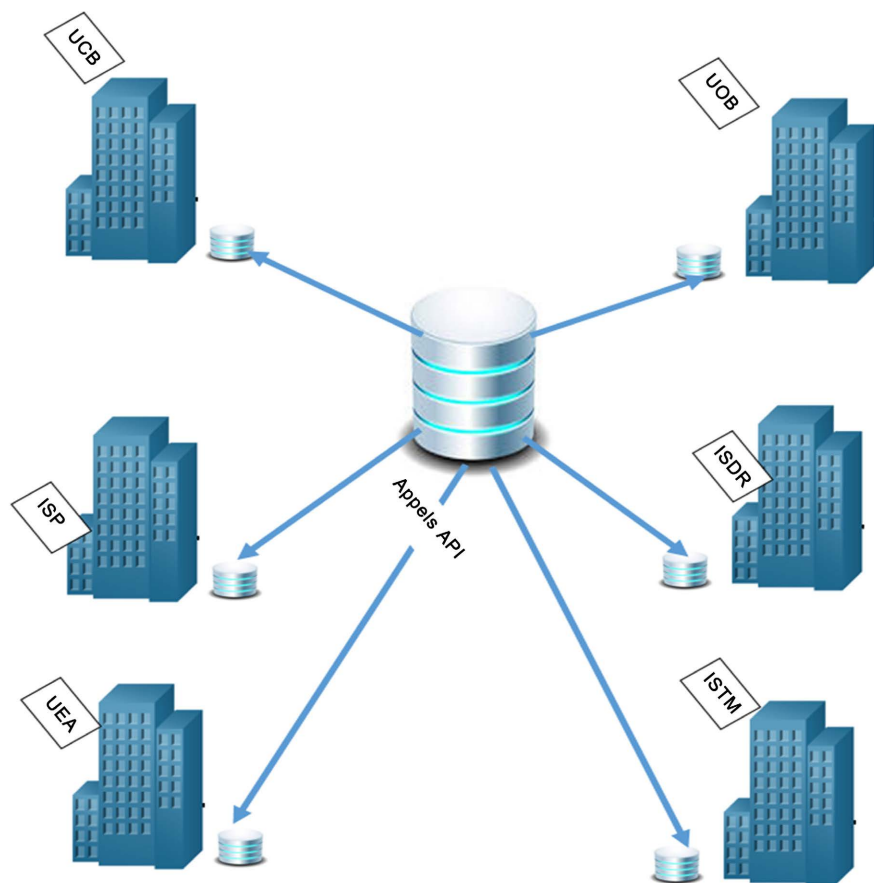


Figure 3. Physical architecture.

Furthermore, from a technical point of view, the deployment of this system would require the standardization of operational systems and the acquisition of a server machine at different universities and colleges, a central server machine and finally an Internet connection for data sharing with the latter, as illustrated in **Figure 3**.

6. Conclusions

Universities, particularly those in South Kivu province, handle a wide range of student data in disparate files. This makes it difficult to get an overview of the

data for analysis purposes.

In this study, we set out to develop these operational information systems into a data warehouse system. The role of the data warehouse will be to collect, organize and archive data from a variety of sources on the curriculum of students at all the province's universities. One of the benefits that both the Ministry of Higher Education and the universities will derive from this tool will be the creation of a model that can guide South Kivu students in their choice of option and university.

We analyzed all data sources concerning the information systems used in the six major universities of the Province. Applying the bottom-up approach to warehouse design, we designed our data warehouse model to store data on a student's academic career.

The present study is limited to proposing the architecture of the data organization from a conceptual point of view and implementing a first prototype. We have just created an information system that will store all the information on students' career paths in all the universities in South Kivu. In addition, we propose the pentaho tool for data collection, integration and transformation, and SQL server for relational data storage. Finally, we have presented the form of data ready for analysis in the form of a data set using SQL queries.

Given that the environment in which this system is to be deployed is a competitive one, a number of organizational and technical constraints need to be highlighted. A charter will have to be drawn up for the use of this data by various third parties, and a security system for access to the data will have to be put in place.

Looking ahead, once the data has been collected and stored in the warehouse, further studies can be carried out to model and understand certain hitherto unexplained phenomena, using data mining and machine learning algorithms.

Conflicts of Interest

The authors declare no conflict of interest regarding the publication of this document.

References

- [1] Inmon, W.H. (2005) Building the Data Warehouse. 4th Edition, Wiley, New York City.
- [2] Matthieu, L. (2014) Entrepôt de données: De l'alimentation des données au décisionnel de l'établissement. p. 128.
- [3] Kimball, R. and Ross, M. (2022) The Data Warehouse Toolkit. The Definitive Guide to Dimensional Modeling. Ralph Kimball, Margy Ross.
<https://www.decitre.fr/livres/the-data-warehouse-toolkit-9781118530801.html>
- [4] Desnos, J.-F. (2001) Agence de Modernisation des Universités et Etablissements. p. 22.
- [5] Kimball, R. and Ross, M. (2003) Entrepôts de données: Guide pratique de modélisation dimensionnelle. 2nd éd, Vuibert informatique, Paris.

- [6] Teste, O. (2000) Modélisation et manipulation d'entrepôts de données complexes et historisées. Ph.D. Thesis, Université Paul Sabatier-Toulouse III.
- [7] Trinidad, S.E.M. (2005) Entrepôts de données pour l'aide à la décision médicale: Conception et expérimentation. p. 165.
- [8] Kimball, R. (2000) Concevoir et déployer un data warehouse. Eyrolles, Paris.
- [9] Jensen, C.S., Pedersen, T.B. and Thomsen, C. (2010) Introduction to Data Warehousing and Business Intelligence. p. 72.
- [10] Sukadi Mangwa, C. (2019) Modernisation et bonne gouvernance universitaire: Implémentation d'un logiciel de gestion académique au sein de l'Université de Lubumbashi en R.D. Congo. présenté à Le "côté obscur" de la communication des organisations. <https://dial.uclouvain.be/pr/boreal/object/boreal:242653>