

A Data Warehouse-based Mining System for Academic Resource Capacity Planning

Wilson Nwankwo

Ph D, MCPN, PMP, RMP

Senior Lecturer, Systems & Security,

SouthEastern College of Computer Engineering & Information Technology
Owerri Nigeria

Abstract—Academic resource capacity planning is fundamental to the success of every academic program in any institution whether at the primary, secondary or tertiary level. As a matter of fact, academic capacity planning is a continuous process often involving both core academic and non-academic personnel of an Institution. As simple as it may sound, academic capacity planning may make or mar the quality of students trained as well as the reputation of the Institution amongst its peers. This is why it is vital for every academic planner particularly at the department level to carefully consider important indicators such as: course requirements, previous student performances, academic achievements, future goals, etc. during the planning process. This research studies the fundamental processes in selected Institutions of Higher Learning in Nigeria as it affects academic planning specifically during the allocation of courses to academics during each semester and develops a model program based on data mining of historical performances of academics and students. The model when implemented would provide the necessary infrastructure for drawing insights through predictions on the results of mined data thereby enhance decision-making as to the allocation of appropriate courses to suitable academic resources.

Keywords— Data Warehousing; Academic resource; Capacity planning; Data mining; Predictive Analytics; Higher learning

I. INTRODUCTION

Planning may be described as an intelligent process often involving thinking, reasoning, and organizing of tasks around certain specific objectives in a bid to achieving goals. In the academic sphere, planning is sine qua non to the growth and development of the human and material resources in the Institution. Academic planning encompasses planning and approval for new academic programs (degrees, majors and certificates), substantial changes to those programs (renaming, mergers, suspending admissions, discontinuations), planning and approval for academic departments and centers/institutes, working with faculty governance on the development of policies that support academic planning processes and changes associated with Subjects and courses[1]. Failure to plan always imply acceptance of failure during implementation or an implicit plan to fail.

Academic resource capacity planning is central and vital to every academic planning process as it influences to no small measure the effectiveness of learning and teaching. Capacity planning is a thorough and very important process usually requiring a professional with unique skills in private businesses. It is usually geared engaging the best hands that will optimize productivity. It is also employed as a means of efficient and effective resource management by private

businesses [2]. Academic resource planning though may not be geared specifically at profit maximization, is aimed at engaging the best resources for teaching of approved academic courses and by so doing enhance productivity. Productivity in this spectrum is not measured by profitability in terms of monetary value but the extent and quality of knowledge and skill transferred to the learner by the teacher (academic resource). This may be determined by examining the various performances of the students being taught over a period of time.

Academic resource capacity planning enables academic decision-makers to commit high priority courses and subjects with confidence to skilled academics within the Institution, that have demonstrated experience and excellence in the areas in question. Thus, resource capacity planning provides decision-makers with the:

- Ability to determine academic resource availability and course requirements and perform a matching of performance needs against a given academic resource
- Insight on resource need and performance gaps
- Flexibility of performing what-if scenario analysis to find the optimal solution set.
- Ability to create resource projections to resolve future resource imbalances

Thus academic resource planning may be considered a process of merging and balancing academic resource capacity against some concrete course performance requirements in order to achieve and sustain optimum learning and skill transfer. This optimality is to be measured continually by the performance of students.

Data warehousing and mining have grown to be the technology of choice in any organization where large volumes of data are imminent and where timely and accurate decision making is based on available data. Data warehousing is aimed at the timely delivery of the right information to the right individuals in an organization by harnessing concrete data from disparate data sources such as legacy files and databases. Data warehousing is a pragmatic and revolutionary means for supporting the analysis and monitoring of critical processes [3]. Like in many other areas, data warehousing plays an important information consolidation base and constitutes a key component of modern organizational intelligence and analytics systems. Generally, academic institutions generates data in various areas such as: students' records, human resources, projects, accounting and payroll, budgets, examinations, academic profiles, library information,

etc. and it is not unusual to see large file cabinets in many offices in such institutions whether public or private. Data mining is an evolving science as well as an art.

A. Problem Definition

In simple terms, every policy on education revolves much around manpower development that is, equipping students in academic institutions with appropriate knowledge and skills for growth and socioeconomic advancement. Policies do not translate to goals on the surface but require careful implementation. Prior to implementing any academic policy, there must be well thought-out plans to guide the policy implementation process. The bane of academic planning is to enhance coordination with resultant growth and sustenance of quality academic performance that is in consonance with regulatory standards and international best practices in education. Academic resource planning is a critical aspect of academic planning. In most institutions in Nigeria and sub-Saharan Africa, academic resource planning is a tasking and cyclical process particularly in public institutions where sociocultural and religious affiliations exert significant influence on procedures and practices regardless of laid down rules. This is worsened by inadequate manpower in most institutions. Even where the number of academics available is relatively sufficient, the question as to which academic personnel would meet the average performance requirements set for a particular course or subject becomes an associative prevailing factor. This is often a challenge to many heads of academic departments who usually chair the course allocation panel or board. The "HOD" as he/she is fondly called, in many occasions allocates courses randomly regardless of course requirements, current course achievements, previous students' performances and future expectations. The challenge may even be more overwhelming in institutions like Universities, which run different academic programs such as undergraduate and graduate programs; amidst various crippling segmented politicalized academic cliques (in various departments) with varying group interests which often hijack formal procedures in a way that do not in any way benefit the average student.

As the allocation of courses to academic resources is a done on semester by semester basis, the head of department is burdened by two key problems:

- How to ascertain courses or subjects which do not have adequate academic resources
- How to gain insight as to determining an academic resource to undertake a course
-

Conventionally, these two problems may require access to physical records of the available academic resources. These problems are in no doubt beyond the records whether physical or electronic because adequacy is not reflected by the number of available academic resources but by: the previous and present student performances, and the future expectations. In addition, gaining an insight as to matching a course against an academic resource is not possible without a deep research into historical data on the department. The question as to how to simplify this process of gaining insight is the problem underlying this study as well as the basis for its conception.

B. Objectives of the Study

The objectives of this study are:

- a. To study the basic academic resource capacity planning processes vis a vis the basis for the semester allocation or allotment of taught academic courses to the various academic staff in institutions of higher learning;
- b. To apply the principles of data warehousing in creating a central repository that would concretize historical records in areas that are relevant to academic resource planning;
- c. To integrate data mining techniques that use statistical models to make predictions thus enhancing effective allocation of academic courses to most proficient academic resource based on a number of prevailing factors on a periodic basis.

C. Scope of the Study

This study is restricted to academic resource capacity planning in institutions of higher learning in the sub-Saharan Africa with Nigeria in focus. Academic resource capacity planning is a multifaceted task chain which spans through many units or departments with administrative functions. Typical academic resource capacity planning tasks may include:

- a. Design and review of academic resource policies in line with global and national standards;
- b. Ensuring compliance with fiscal and human resources policies and procedures;
- c. Planning academic and non-academic resource hiring;
- d. Development of new financial and human resources systems;
- e. Determination and review of Academic resource capacity needs of various academic departments;
- f. Matching every academic course specification to available academic resource proficiencies with emphasis on previous performances;
- g. Preparing and monitoring the annual budget;
- h. Developing annual enrollment targets;
- i. Managing contingency funds and ensuring fiscal solvency of departments;
- j. Preparing financial reports for and managing the accounts of units within Academic Affairs Administration;
- k. Conduct facility planning and coordinate academic space.

In this study we examine how the challenges encountered by decision makers in the academic circle as to harnessing and extracting useful information from large volume of historical data generated by academic departments in the process of ensuring the attainment of academic successes, are eliminated through the use of mining techniques applied to a data warehouse. One key area of emphasis is the matching of every taught academic course to academic resource persons based on academic proficiencies and previous performances.

D. Significance of the Study

Data mining and Data warehousing are two computing technologies that are highly associated with analytics and intelligence gathering thereby helping decision-makers realize effective decisions at all levels. This study presents a pragmatic and implementable prototype of a system for not just harmonizing the data generated from the academic resource allocation processes and consequent results into a data warehouse but a computer-based model for extracting and mining hidden information for use in making predictions. Academic resource allocation decision makers need to be empowered with the vital tools to make important demand commitment decisions based upon limited resources or capacity information. In a nutshell, this study is important in that it will:

- a. Support the mining and analysis of academic staff data, to determine their relationships with staff performance;
- b. Support the identification of areas where academic staff performances are not improving, and thus enhance decisions-making on resource allocation reviews;
- c. Provide a simple environment for data analysis and reporting;

II. METHODOLOGY

Developing a new system is a problem solving process and requires the selection of a coherent method or suite of methods often called a methodology. Methodology may be described as a framework used to structure, plan, and control the process of developing an information system and consists of steps, methods, techniques and procedures which govern the collection, analysis and design of a particular project [3]. Some of the common methods that are employed to develop data warehouses include: Top-down, Bottom-up, Agile, and Object-based methods. The object-based method was employed in this study due to its support for formal analysis, modeling of dynamic complex systems, and rapid development and maintenance. In

A. Documentation

This study involved an investigation and documentation on three institutions of higher learning drawn from the South-East, South-South and South-West geopolitical zones of Nigeria respectively. That is, one institution from each zone. These institutions include: Federal university of Technology Owerri in the South-East, University of Lagos in the South-West and Rivers State University of Science and Technology Port Harcourt in the South-South zones respectively. Relevant data including sample historical data were collected from the Departmental heads of two academic in each institution.

In each institution, Academic resource capacity planning as it relates allocation of taught courses in each department is done based on the courses offered by the given department with particular emphasis on the available academic resource capacity. The process may take note of the following factors:

- a. Requisite academic qualifications;
- b. Publications;
- c. Previous course handling experience;
- d. Completed researches/Areas of specialization;
- e. Course performance histories.

B. Development tools

The tools that were used for analysis and development are: Erwin Data modeler; Microsoft Visual Studio 2012; Pentaho Data Integration (Kettle), and Microsoft SQL Server 2014.

C. Analysis of the Present System

To match a course with an academic resource does not seem to be an easy task when the factors listed above are carefully considered by a decision maker. However, to present the workings of the existing system, we specified the following:

- a. The actors i.e. those who play a role(s) in the system;
- b. The process model that is, what happens in the system or what is going to happen in the new system using an activity diagram;
- c. External users(if any) that is, individuals outside the logical boundary of the business area who may use the system;
- d. Use cases that represent what the participants are doing in the system/what the users will do with the new system;
- e. The interaction among two or more classes or objects using sequence/collaboration diagrams;
- f. Classes of objects/entities, their attributes, relationships and methods using class diagrams

Actors: academic staff, head of academic department, course evaluation officer, administrative officer, course allocation officer.

The process model of the existing system is presented using the activity diagram in Figure 1. The process model is summarized as follows:

- a. Administrative officer/Secretary compiles the list of courses offered in a semester and presents in a single spreadsheet document;
- b. Head of Department reviews the academic resource capacity in the department visa a vis core and non-core departmental courses offered in the department;
- c. HOD/Course evaluation officer evaluates each academic staff profile including research areas, publications, qualifications, ranks, previous course handling experience;
- d. HOD assigns courses based on available academic resource while considering academic profiles and previous experience;
- e. Course allocation report is tabulated, printed and sent to the academic staff.

D. Use case analysis

A use case is an abstract representation of a functionality that users need from the system. In object-oriented analysis, use cases are also used to depict the requirements analysis process. The functionalities defined by a use case are represented using the use case diagram. Figure 2 is the use case diagram representing all the use cases associated with the present system of matching available academic resources to courses. The interrelationships between the use cases are also established.

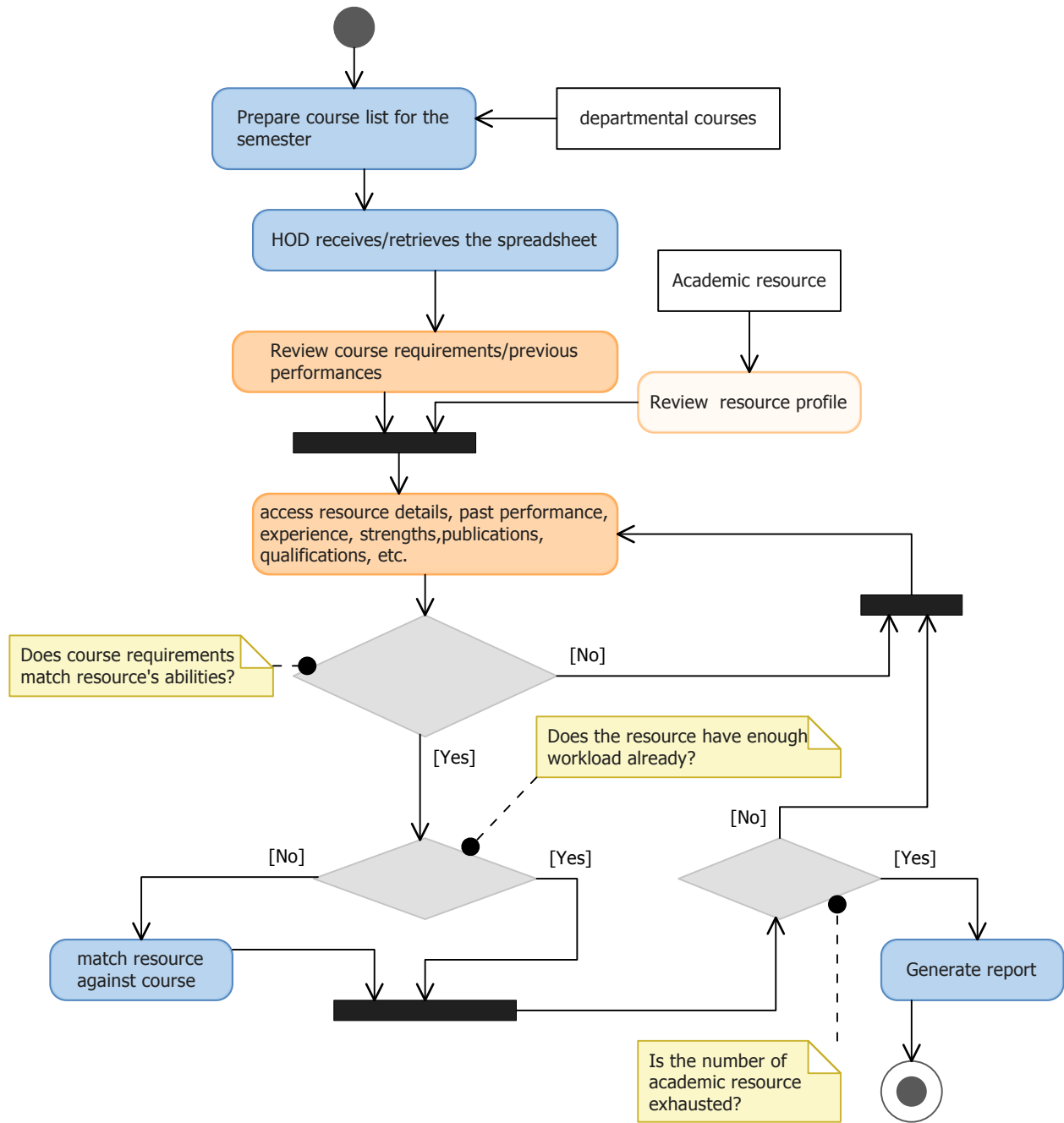


Fig. 1. Activity diagram representing the process model of the existing system

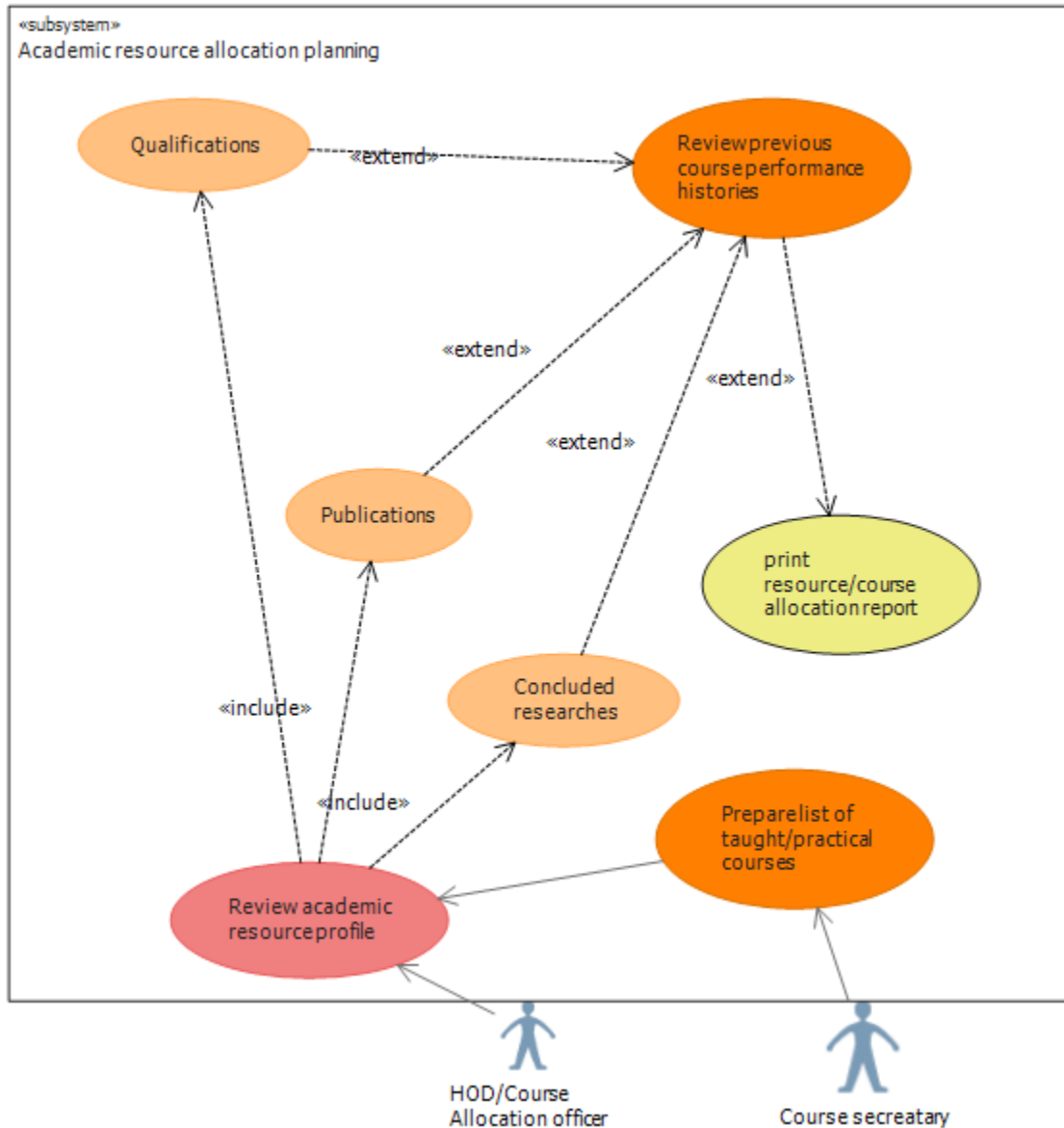


Fig. 2. Use case diagram of the Academic resource to course matching process

E. Domain analysis

In this phase, we identified and defined the objects/concepts inherently present in the use cases using a conceptual model. The model described what data/information is (would be) managed by the course matching process chain, and what data flow between users and the system. We used class diagrams and the unified modeling language for the conceptual modeling. The data represented by a class is broken into: concept and association. The concept (condensed form of an object/class) is the representation of complex

information that has a coherent meaning in the scanning operations domain. Concepts aggregate attributes and may be associated to each other. The identified concepts in the existing are presented in the model in Figure 3. The conceptual model consists of condensed classes with associated relationships. The arrows show the relationships. The dotted arrow shows a dependency relationship, a solid arrow shows an association whereas an arrow with a triangular pointer shows inheritance.

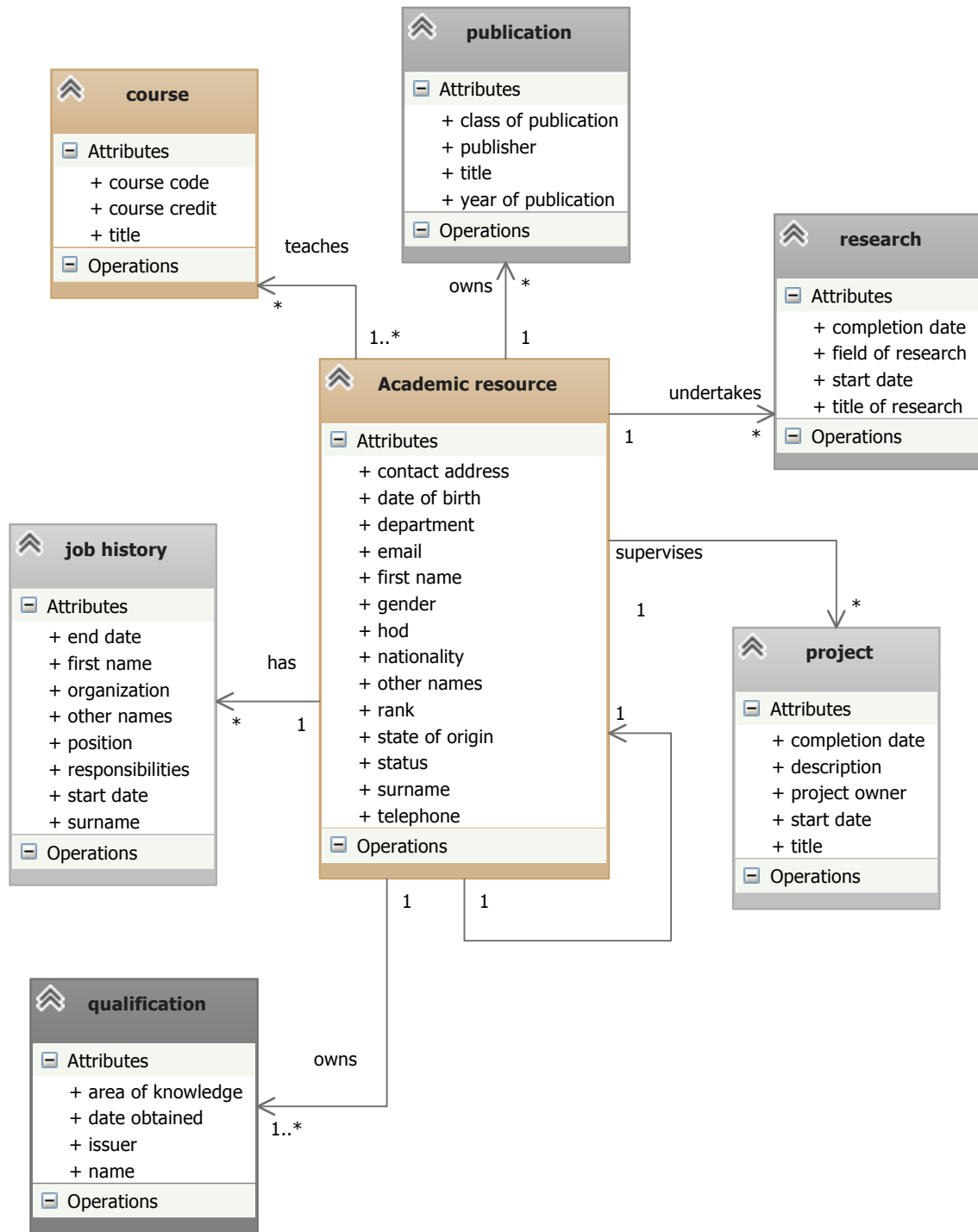


Fig. 3. Conceptual model of the central objects of the existing system

Identifying classes and their relationships is a very prior to implementing the requirements of any system. The object schema represented by the class diagram in Fig. 4 shows the various classes that form the foundation of the proposed model. Each object/class has static (attributes) and dynamic

(behaviour/methods) characteristics. We are concerned with the attributes alone, as shown in the model in Fig. 4. In the model, the primary (PK) and foreign key (FK) attributes have been appropriately identified.

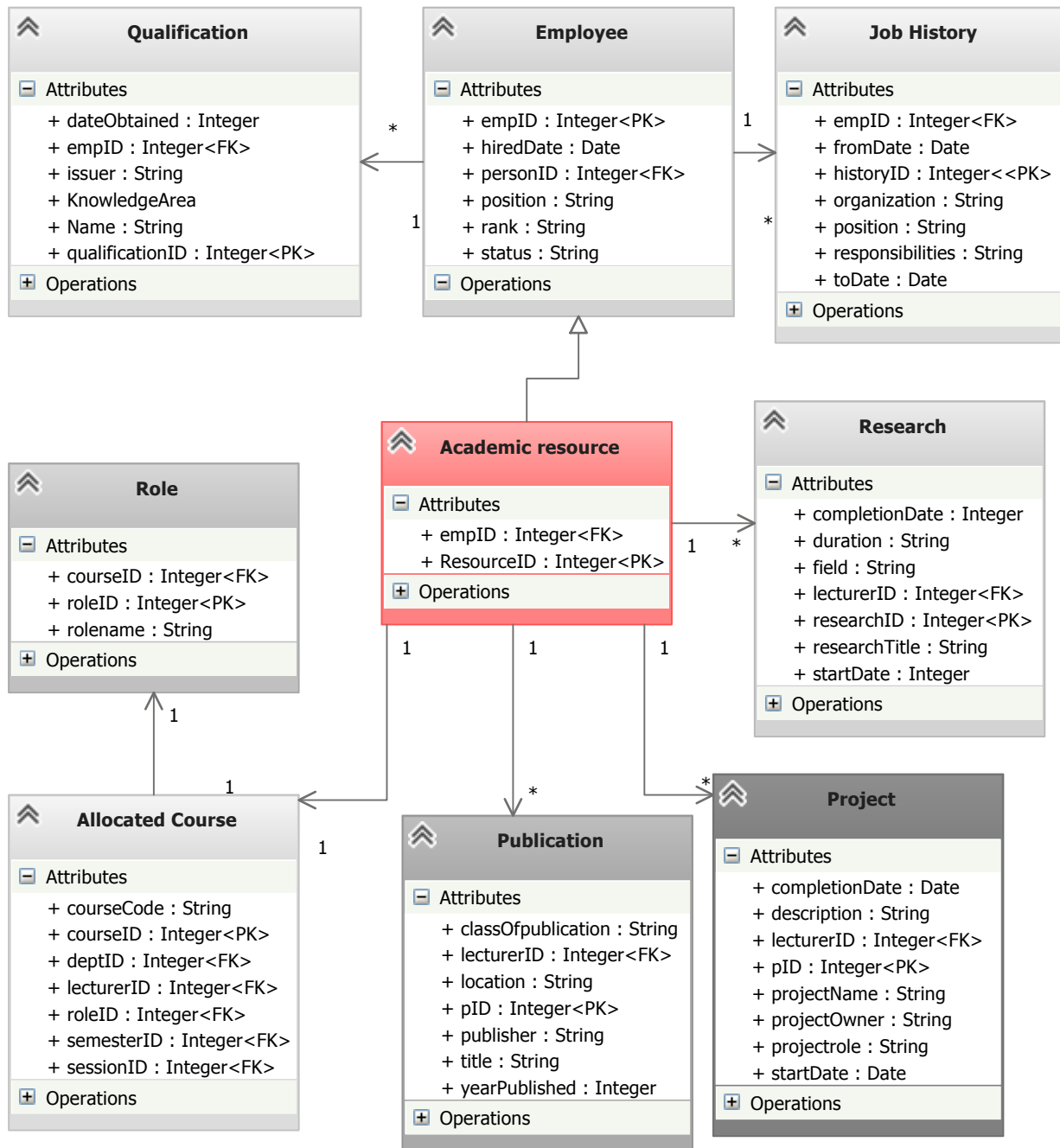


Fig. 4. Class diagram of the existing system

F. Problems in the existing system

The existing system of performing course-matching is fraught with complexities as itemized below:

- a. The existing system of resource to course matching often adopts a rule of thumb without recourse to previous performances and future expectations but on availability or which academic staff has been taking the course from time immemorial;
- b. The process requires a lot of filtering and sorting by administrative staff or the decision-maker and in most cases become cumbersome with increasing number of courses and faculty members;

- c. Generating intelligence reports are often slow, ineffective and inefficient because the decision-maker must spend time to access all the various records needed to evaluate the suitability of the academic resource as regards the course in question.

These challenges were further validated with the use case analysis of the proposed model below.

G. Use case analysis of the proposed system

The functionality expected in the proposed system is reflected using the use case diagram in Figure 5. Here, the improvement is seen in the integration of warehousing to extract relevant data and the subsequent use of mining techniques such as clustering to establish the most suitable resource for the given course based on the rated parameters.

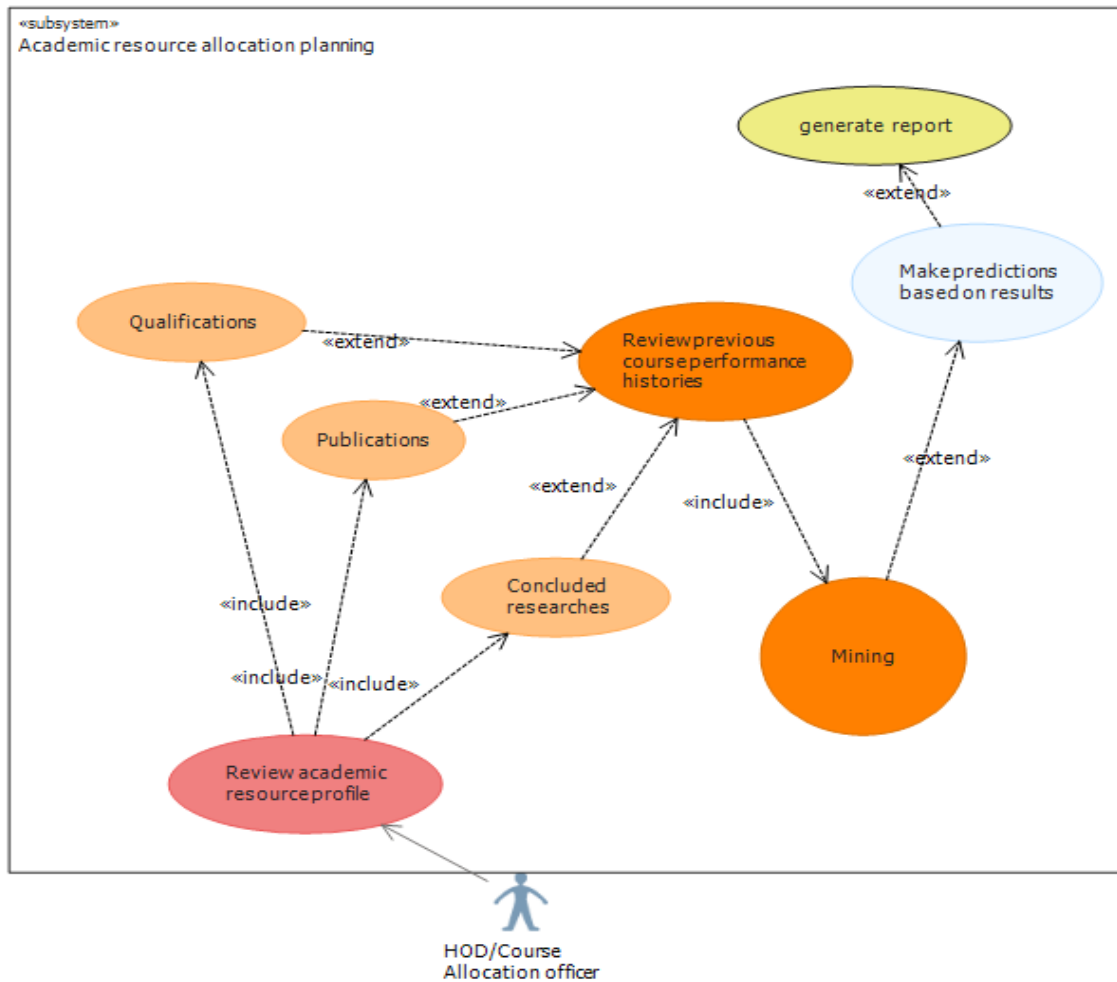


Fig. 5. Use case diagram of the proposed system

III. THE DATA WAREHOUSE

The prototype design is restricted to the vital components of the data warehouse that would be useful for mining for decision making. The warehouse data is synthesized from data sources such as spreadsheet and database files. The design consists of three phases namely: conceptual, logical, and physical. We employed the Data mart approach. A data mart represents a unit or departmental process within an organization. It is composed of a fact table and associated dimension tables. The combination of one or more data marts constitutes the data warehouse. Table 1 shows the fact and dimensions in the in the data mart.

TABLE 1 ACADEMIC RESOURCE PACKAGE

Academic Resource package	Dimension	Fact
	Academic	FACTFACULTY
	Course Allocation	
	Qualification	
	Publication	
	Research	
	Project	
	Feedback	

Building an Academic resource data warehouse to capture all functional warehouse needs would be a tasking project; hence the data mart offered more flexibility in that a set of recognized data analysis requirements for a unit or department could be captured on demand using a data mart. To this end, we have a single data mart called the academic resource schema. The schema is designed using package stereotypes. Figure 6 shows a conceptual model containing packages. In figure 6, a package contains fact or dimensional class. Thus, the packages constitutes a single star schema (data mart).

A. Conceptual model design

What is important in data warehouse design is not all the data in a transactional data store, but those elements, which are the essential drivers of the intelligent decision-making process. These essential drivers are the grains which grouped into a single package called the 'scanning schema'. The schema is divided into two parts: fact and dimension classes. A fact represents measures and context data. A dimension is a set of data that describe one business dimension. Dimensions determine the contextual background for the facts. Both parts are derived from the classes in Fig. 4 above. Figure 6 shows the conceptual star schema.

B. Logical model design

The logical schema is an extension of the conceptual schema represented earlier using the stereotypes in Fig. 6. We established the attributes of each class, and the relationships (using association) between the fact class and the dimension classes. Thus, the fact class is associated to the dimension

classes. Each class (fact/dimension) is identified by a unique object identifier {OID} attribute. Figure 7 shows the logical data model of the data warehouse. Every dimension class name is preceded by a “Dim” prefix likewise a fact class name is preceded by a “fact” prefix.

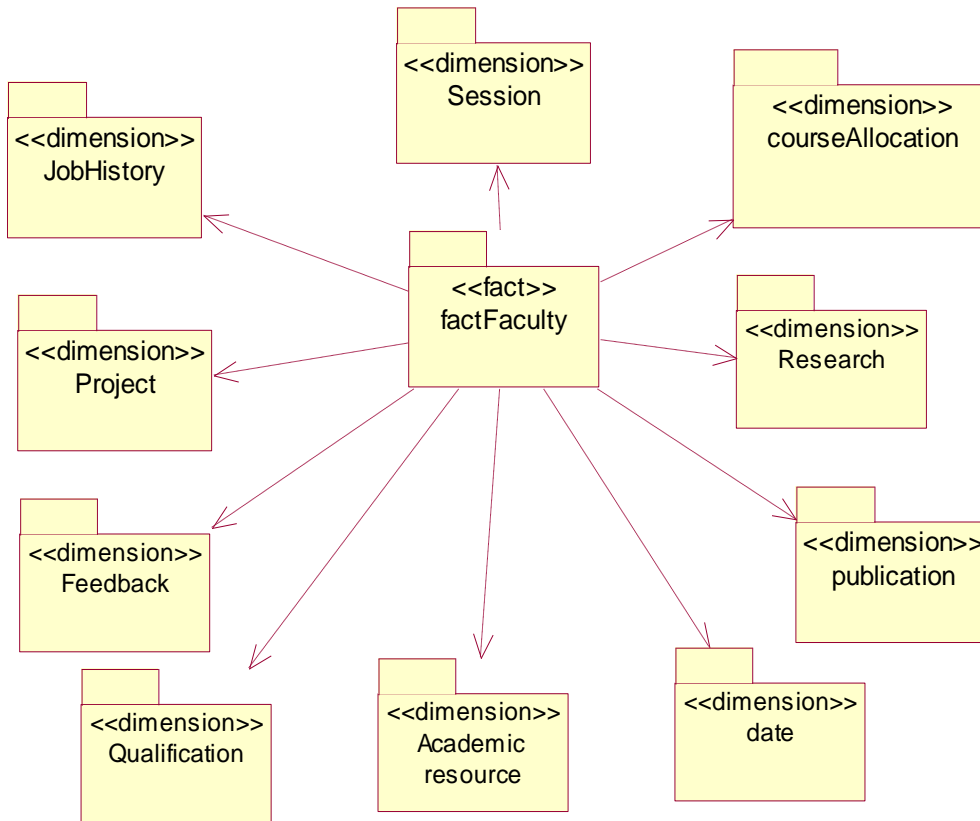
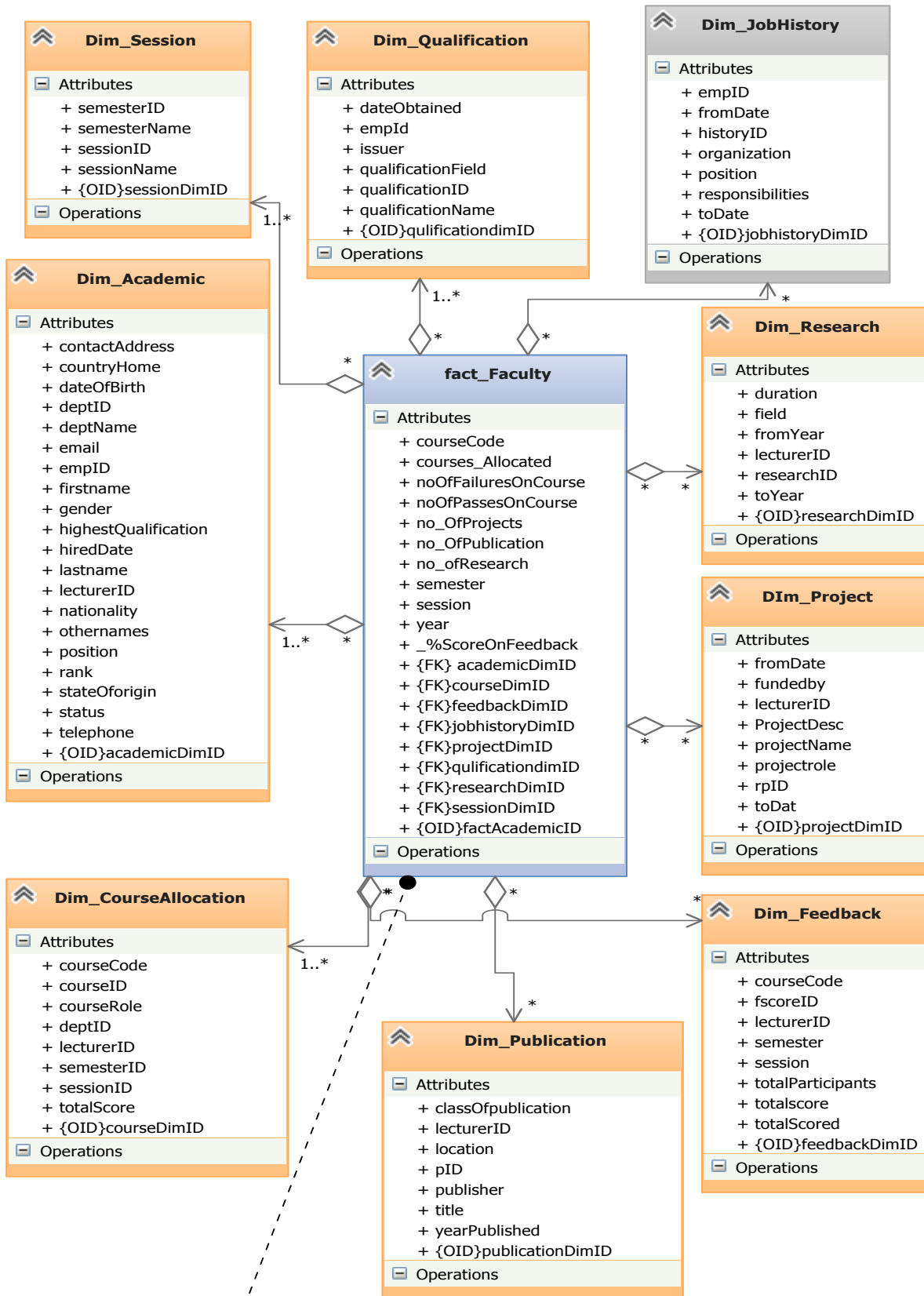


Fig.6. Conceptual star schema



$\%ScoreOnFeedback = (DIM_FEEDBACK.totalScore / DIM_FEEDBACK.totalScore) / 100.$
 It is the total percentage feedback score of the Academic resource on an allocated course in a semester. This score can be compared to the pass and failure rates on the course respectively and data mining can be used to relay connections between these parameters

Fig. 7. Logical model of the Data warehouse

C. Physical design

The most important activity in this phase is the conversion of the logical design into a physical model by using database system structures such as tables, tablespaces, etc. The various classes in the logical model in Fig. 7 would be mapped to tables, relationships to foreign key constraints, attributes to fields, and object identifiers to primary key constraints. Further denormalization may be necessary since the emphasis is on query performance in a data warehouse other than storage optimization. The data type, field length, primary and foreign key constraints were all defined. Due to the fact that some attributes of one or more dimensional tables may be updated in the course of the business, we used slowly changing dimensions (SCD). Though many types of SCDs exist, we identified only three which are relevant in this project namely:

- a. SCD – Type 1, in which the old data in the record is overwritten with new data during updates;
- b. SCD–Type 2, which creates additional record with the new data at the time of change and can track changes in the history of the data but required a generalized key to record all the iterations of the original record;
- c. SCD–Type 3 creates new fields in the record for the new data and the time of the change, and tracks original and current values only, losing intermediate values.

The SCD type 2 is widely used in this study. We employed the component diagram in Figure 9 to show the overall physical design of the data warehouse. The flat files stored in a designated network directory are fed to the ETL program which loads the files to the scanning staging database tables. The ETL also loads the data warehouse tables. The fact and dimensional tables are stored in different Tablespaces for optimal performance.

D. Extract-Transform-Load (ETL) design

The ETL process is responsible for extracting, transforming into a consistent form. The process is composed of six tasks:

- a. Selecting the sources for extraction;
- b. Transforming the sources, in which the data is transformed into new data by filtering data, converting codes, performing table lookups, calculating derived values, transforming between different data formats, automatic generation of sequence numbers (surrogate keys),etc.;
- c. Joining the data sources;
- d. Selecting the target to load;
- e. Mapping source attributes to target attributes;
- f. Loading the data from the sources.
- g.

The ETL process runs periodically (via a script) to automatically extract, transform and load data to the data warehouse database. Data is extracted from the flat file (Microsoft Excel or compatible file) into a staging database (a temporary data store) from where data is ultimately loaded into the data warehouse tables. The data warehouse would be accessible by all intended users irrespective their locations.

This is easily possible through the internet or over a virtual private network.

IV. MINING

Data mining has been defined by researchers as a process of searching hidden information that can be turned into knowledge thus providing information for strategic decision making or solving a fundamental research question. There are many techniques, algorithms and models used in mining data. Here, we employed the Modified Naïve Bayes algorithm in building the mining model. Naïve Bayes is a classification technique. Classification techniques are used to discriminate data according to their classes, and generalize the findings in order to explain a phenomenon. The Naive Bayes algorithm is based on Bayes theorem, and used in predictive modeling. The algorithm uses Bayesian techniques but does not take into account the dependencies that may exist. It is less computationally intense and very useful for quickly generating mining models to discover relationships between input columns and predictable columns. The algorithm could be used to conduct an initial data exploration, and thereafter, apply the results to create additional mining models with other algorithms. It calculates the probability of every state of each input column, given each possible state of the predictable column.

V. IMPLEMENTING DATA WAREHOUSE AND MINING

A. Requirements

Hardware: At least a 2.4 GHz quad-core network-ready computer with 8GB RAM, running Microsoft windows 7/8 OS or Linux and forming part of a computer network with an active file server where the spreadsheet workbook or text-compatible files could be localized centrally.

Software: Microsoft SQL Server Database 2014(32/64-bit) or higher and Pentaho Data integration (PDI) 5.2 or higher; Java development kit (JDK) 1.7 or higher

The implementation phase is divided into six phases:

- a. Configuration of a network directory in the local network where the source files (spreadsheet/text files are located);
- b. Installation and configuration of MS SQL Server database, Pentaho Data Integration (used to design the extract-transform-load program);
- c. Creation of the ETL data transformation schema and associated jobs;
- d. Creation of a physical staging database tables where data are extracted and loaded to by the ETL program;
- e. Creation of the data warehouse tables (dimensions/fact);
- f. Loading of staging tables and data warehouse tables.

B. Loading the staging and the data warehouse tables

The staging tables are loaded from the transaction file sources (spreadsheets particularly the Microsoft Excel workbook comprising of many worksheets). This study used the community edition of Pentaho Data Integration (PDI) and Microsoft SQL Database Management System for testing the model. PDI is an open source tool and do not require expensive hardware to run. The PDI software is used to

develop an ETL model which maps each worksheet in the Excel workbook file (located on a network file system) to a specific table in the staging database (which may be remotely located). Transformation and loading to the dimensional tables of the data warehouse are made from the staging tables using scripts generated by the ETL model.

C. Working with PDI

PDI (community edition) has four components:

- Spoon - a single user graphics platform used to create/build ETL jobs and transformations;
- Kitchen – a standalone command line component used to execute jobs created with spoon;
- Pan – a command line component used to execute transformations and jobs created with spoon
- Carte – a small web container which can be used to set up a dedicated remote ETL server (Data integration server).

PDI 5.2 is java-based software and requires JDK 1.7 or higher to run; PDI does not come with a Microsoft SQL Server driver by default; PDI cannot communicate with MySQL by default; We need a Microsoft SQL Server connector (a separate jar file) to enable Communication between PDI and Microsoft SQL Server; and Microsoft SQL Server connector/JJDBC driver for Microsoft SQL Server) could be downloaded from <https://msdn.microsoft.com/en-us/sqlserver/aa937724.aspx?f=255&MSPPErr=2147217396>. An Open-Source MS SQL Driver (JTDS) can also be used to connect to MS SQL Server. JTDS is an alternative, non-vendor-supported driver. To install, the driver is unzipped and placed in the 'lib' sub-directory in the Pentaho directory. For instance on windows machine we copy to the following sub-directories:

DI Server: C:\pentaho\server\data-integration-server\tomcat\webapps\pentaho-di\WEB-INF\lib/

Spoon: data-integration/lib

The server and spoon must be restarted after a successful copy.

We install JDK and JRE (if not done already) and set the following environment variables: CLASSPATH, JAVA_HOME and JRE_HOME.

The *classpath* is set to include the location of the JDK on the system e.g. “.;C:\Program Files\Java\jdk1.7.0_71”; JAVA_HOME is set to point to the JDK directory and JRE_HOME is set to point to the JRE directory

D. Building Extract-Transform-Load with spoon

We divide the ETL construction into two stages: Building the transformations and Building the jobs. Building the transformations involves three levels in this phase:

- Staging transformation
- Data warehouse dimension transformation
- Cube/fact transformation

Prior to the building of transformations and jobs, we need to establish a repository for storing/tracking transformation/job meta data. A repository is established through the Tools menu.

A new transformation must be created in spoon before the database connection is visible under the view tree. A new

transformation could be created either from the file menu or the view tree. At least two database connections are necessary and may be labeled ‘*STAGING*’ for the staging database and *AcademicDW* for the data warehouse database (see Figure 10) The staging transformation consists of steps (transformation elements) and hops that interface with the Excel workbook; to extract, transform and load data from the workbook into the staging table. The actions taken are:

- Insert or put Microsoft Excel input (from the design tree) in staging transformation window
- Select the worksheet(s) to map to a staging table;
- Set the content types and output fields characteristics
- Add some transformation steps
- Map the results to a database table output step(staging table)

F. Creating the Mining project in SQL Server

In Microsoft SQL Server, the Naïve Bayes algorithm has a set of unique parameters which may be altered to achieve intended results in a mining problem. We implemented this algorithm using a data analysis project tagged *HIGHER* in Microsoft Visual Studio. Data sources, data source views, data mining structures, and data mining models were added. The data source and data source views usually interface with the data that would be used for mining. Simply put, the data source is a connection string that indicates the data location; whereas the data source view (DSV) is an abstraction layer that enables a developer to modify the behaviour of a data source when required. The data warehouse is used as the data source. A mining structure defines the domain of a mining problem, whereas a mining model is the application of a mining algorithm to the data in a mining structure. A mining structure contains a list of columns that have data and content types, bindings to the data source, the mining model, and some optional flags. The flags control how the data is modeled. The specification of a mining model contains an algorithm with its associated parameters, plus a list of columns from the mining structure. Each model in a structure can use a different algorithm or the same algorithm with different parameters, and/or a different subset of the columns in the structure. A function and algorithm-specific modeling flags could be assigned to each column/field in a model. These features of a model allow easy testing of different hypotheses on the same data set.

G. The Mining Problem

The mining problem is How to discover an academic resource that is better placed to handle a taught course in a given semester by using historical performance, academic staff and summarized course result data. The data requirements for a Naive Bayes model are:

- A single key column: this must contain only one numeric or text column that uniquely identifies each record;
- Input columns: all of which must be either discrete or discretized, are important to ensure that the input attributes are independent of each other, particularly when the model is to be deployed for prediction; the reason being that, if two closely related columns of

data are used, the effect would be to multiply the influence of those columns, which can obscure other factors that influence the outcome.

treated as an input; this is useful when exploring a new dataset, to find relationships among the columns.

- c. At least one predictable column, which must contain a discrete or a discretized value, which can be

The faculty fact table (fact_Faculty) is used as the source data table and a subset of the data is selected, which include the fields defined in table II below.

TABLE II: DATASET SPECIFICATION FOR NAÏVE BAYES ALGORITHM

Variable	Data type	Content	predict	input	function
FactAcademicID	long	Key	No	no	Faculty Fact table ID that identified each row in the fact table
Course code	text	Discrete	no	yes	Field representing a unique course
Lecturer name	text	Discrete	yes	no	Represents the full name of the lecturer
LecturerID	long	discretized	yes	no	The ID of a lecturer in a department
_%ScoreOnFeedback	double	discretized	no	yes	The aggregated percentage score computed from all responses from students that took a course in a particular semester
NoOfFailuresOnCourse	long	discretized	No	yes	Aggregated number of students who failed a course in a particular semester
NoOfPassesOnCourse	long	discretized	No	yes	Aggregated number of students who passed a course in a given semester

VI. TESTING AND RESULTS

The course allocation and resource matching rights and privileges are exclusively reserved for a user with a ‘HOD’

role. No other user of the application has access to this function.

The specification for this test case is shown in table III.

TABLE III: THE COURSE-RESOURCE MATCHING TEST CASE

Test case id:	TCCA
Unit to test:	Course allocation module
Role	HOD
Assumptions:	Head of Department assigns course(s) to a faculty staff, after reviewing data mining results from the data warehouse
Test data:	Semester courses, Academic staff/lecturers in the department, prediction results
Flow	<ul style="list-style-type: none"> i. HOD examines list of all courses available at different levels in a given semester; and invoke the data mining module for prediction; ii. Select the mining model prediction tab. By default, the tool is configured to select input data from a table. The default could be changed by clicking the third icon, to switch to a singleton query. See figures 8 below. iii. With the singleton query, supply the input values such as course code, minimum failure rate, pass rate (expected), and a feedback score range. The “lecturerID” which is the field to be predicted is left without a value. iv. In the grid at the bottom of the form, select the existing mining model as the source, and set “lecturerID” as the field. v. Execute the query by clicking the first icon (display results) vi. Go back to the course allocation window; click the ‘allocate’ button, assign course to a lecturer in the course allocation window and commit the transaction to the database
Expected result:	An assigned course is registered against the course lecturer’s profile, and the course status changes on the course list window to ‘allocated’ with the name of the lecturer and semester against the course.
Actual result:	Allocation action succeeds, course is moved from unallocated list to allocated list
Pass/Fail:	Pass

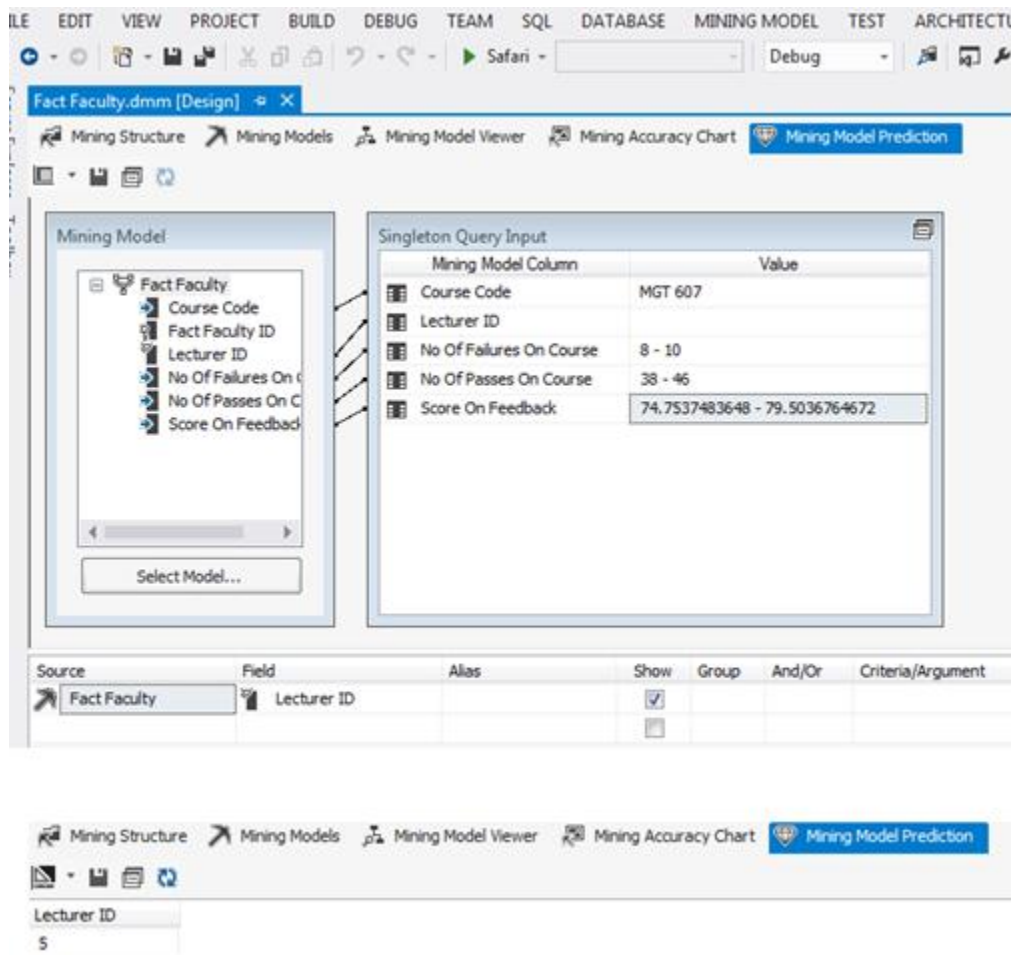


Fig. 8. Academic resource prediction results

VII. CONCLUSION

This study presents a concise model prototype for supporting decisions made by academic resource capacity planners such as the head of department and course evaluation officers of higher learning institutions in Nigeria. The study commenced with a field study of three institutions of higher learning in Nigeria.

The existing technologies and processes were duly observed, documented and reviewed to identify the areas that may need improvement in information consolidation and intelligence decision making as it affects courses-to-resource matching, and subsequently to devise pragmatic solutions. Challenges in the area of data gathering, consolidation, reporting and decision support were prevalent at all three sites. We discovered that there is no advanced technological tool such as that proposed by this study for extracting, consolidating and exploiting data for gainful applications such as in decision making. Considering the large volume of data generated in departmental activities from time to time, the

difficulties experienced by analysts and officers in generating intelligent data for analytic/report purposes are enormous. To improve the situation we proposed and designed a model for implementing an academic data warehouse that would support intelligence gathering, information distribution, reporting, and analysis for decision making.

REFERENCES

- [1] Milner. J(2016) Academic planning, Academic Planning and Institutional Research, University of Wisconsin. <https://apir.wisc.edu/academicplanning.htm>, retrieved 29 July 2016
- [2] Burgess, T.F.: Planning the academic's workload: different approaches to allocating work to university academics. Higher Education. 32, 1, 63-75 (1996).
- [3] Nwankwo, W. & Famuyide, O., "A Model For Implementing Security And Risk Management Data Warehouse For Scanning Operations In Nigeria", International Journal of Engineering Research and Technology, India, April 2016