Business Intelligence Applications for University Decision Makers

Carlo DELL'AQUILA, Francesco DI TRIA, Ezio LEFONS and Filippo TANGORRA
Dipartimento di Informatica
Università di Bari
via Orabona 4, 70125 Bari
ITALY
{dellaquila, lefons, tangorra, francescoditria}@di.uniba.it

Abstract: - The development of a Business Intelligence Application starts with the execution of OLAP queries to extract information from data stored in a Data Warehouse. The results of these queries, together with an opportune data representation, offer a deep synthesis of data and help business users to better discover hidden knowledge than using conventional database queries. Nowadays, information discovery is getting more and more important also in academic environments, because internal evaluation teams have to provide the guidelines to improve the quality of both Didactics and Research. In this paper, we present an approach to implement Business Intelligence technology in the University context. The architecture of the data warehouse and examples of analytic queries for the didactics management are also presented and discussed.

Key-Words: - Data warehouse (DW), Data mart (DM), Online analytical processing (OLAP), Academic application.

1 Introduction

Business Intelligence is an activity based on software technologies whose aim is to support the management and processing of data in an Information System. This data processing, that ends with the production of information usable by managers and business decision makers, starts with collecting and storing large volume of historical and heterogeneous data in a central repository: the data warehouse (DW). For this reason, DWs are the most significant component of strategic decision making for business. In the last years, a new approach to analyze business data has become important for those companies, as banks, financial services, or chains of supermarkets, for which the customer satisfaction is the key of success. In some cases, the increased importance of the Business Intelligence technology has led many companies to review the whole business process [1].

However, in the early years, the costs for the development of a data warehouse were very expensive. Only recently because of the lowering of the cost for developing and maintaining a data warehouse, these databases designed to support managerial decision-making have become functional tools to be used as repository of information [2]–[5].

Also Universities, that until this moment were almost absent in the list of the major users of data warehouses, have accepted to take advantage of developing a decision support system in an academic environment [6]. In fact, nowadays, we can consider the management of a University as critical as the management of a big business company, because the factors affecting an optimal management of a University are the same as those involved in the business processes.

Typical objectives affecting the management of a University are: offering a better quality of the education; managing employees and human resources; managing economic-financial institutions; avoiding wastes.

Several environmental factors encourage academic institutional leaders to adopt an academic data warehouse. These factors include not only decreases in governmental financial support, faculty supplies and research founding, but also increases in student tuition, competition, faculty salaries, faculty support and expectations from students, parents, and employers. Each of these factors generates informational drivers for the development of an academic data warehouse. One driver is represented by the necessity to follow the pace of change affecting business companies; this driver obligates academic institutions to gather information to support strategies and processes that address changes. Another driver is to provide a centralized repository as a centralized tool for all the decision makers to control global resource allocation and use.

Given these information drivers, there are several benefits that can be reached by developing an academic data warehouse [7]–[10]. For example,

1) providing a centralized source of information accessible across different academic units to quickly analyze problems and get satisfactory solutions,
2) supplying the data necessary for developing the Institution's strategic plan, and
3) enabling administrator to timely make better business decisions based on historical data available in different data stores.

This paper presents the architecture and design of an academic data warehouse supporting the decisional and analytical activities relative to the three major components in the university context: didactics, research, and management.

In Section II, we overview the current status of the systems supporting the development of Business Intelligence applications at the national level. Section III describes our Business Intelligence system with regards to the academic architecture. Section IV lists the source databases, and Section V shows the designed data warehouse. Section VI introduces the OLAP layer supporting academic decision makers. In particular, Section VII shows the logical schema of the Didactics data mart. Section VIII presents some typical Business Intelligence applications developed with our system. Section IX reports conclusions and future works.

2 The National Context

In our national context, the most significant experience about developing applications related to the various university management needs has been made by CINECA [11]. CINECA is a non-profit Interuniversity Consortium, made up of 28 Italian universities. Due to its nature, the consortium follows with great attention the national normative evolution, continuously adapting the released applications and/or developing new ones.

The main developed applications aim to manage:

- the legal-economic career and the wage of the academic and technical personnel;
- the function and activities of the employees;
- the career of the students of the Athenaeum and the didactic programming;
- the economic and financial resources.

The consortium proposes also a data warehouse for analytic activities.

Nevertheless, each University adopts only the services it chooses from those developed by the consortium. Moreover, each University has own legacy databases of historical data. It follows the need of data integration in order to access all these information resources mainly for analytic purposes.

Also our University meets the previously described conditions. Therefore, we decided to develop a specific data warehouse integrating all the present and historical data resources.

3 The Business Intelligence Architecture

Business Intelligence consists of applications and technologies that help companies to have a wide knowledge about their own business performances [12]. A Business Intelligence System in the University context has a wide knowledge about the performances on students, teaching staff, and a unique system of analysis and reporting for the supervisory staff of the Athenaeum and for Didactics. A University data warehouse is designed to provide a valid tool that satisfies the following needs:

- the single organizational and administrative structures, such as departments and secretariats for the students;
- a system that supplies in real time data to information external agencies.

Figure 1 shows the University data warehouse architecture structured on the typical multi-level layout.

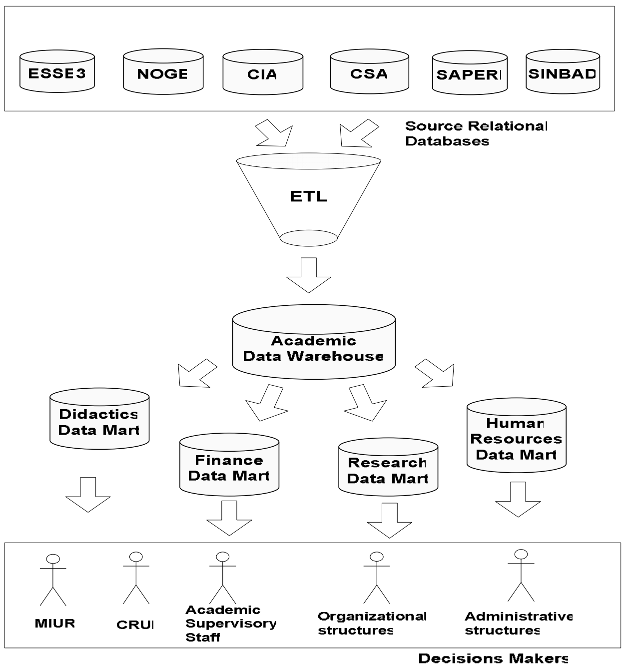

Fig. 1. The academic data warehouse architecture.

4 The Source Databases

Source databases contain transactional data. Figure 1 shows six source databases.

ESSE3 (Secretary and Services for Students) is the actual database that supports all the didactic curricula, and the administrative processes and services to the students with respect for the didactic autonomy of the University.

NOGE (NOt manaGEd) is a secondary legacy database that contains residual historical data about the students enrolled before the ESSE3 introduction.

CIA (Athenaeum Integrated Accounting) is the integrated financial management system that views the University as a business company that distributes specialized services (Research and Didactics, for example).

CSA (Careers and Wages of Athenaeum) takes care of the legal and economic management of the university personnel.

SAPERI is the database of the scientific research competence of the University. It includes also publications and patents of researchers. These data concur, for example, to construct the athenaeum year-book.

SINBAD is the data system for the management of the athenaeum research projects.

5 The Data Warehouse

After the ETL (*Extraction Trasformation Loading*) phase, the data warehouse contains cleaned historical and integrated data.

Given the high complexity of the data warehouse, we are presently concerned in designing and implementing the single data marts that represent the departmental databases. On these, we will then construct the entire data warehouse.

The current data marts allow to model the following academic departmental areas:

a. **Didactics**. This data mart contains data about the career of the students of the Athenaeum. Moreover, there is information on the University formation offer structured in Faculties and Degree Courses.

b. **Finance**. This data mart is developed to run twofold analyses: (a) the analysis of financial documents, and (b) the analysis of general and analytic economic movements.

c. **Research**. The research data mart contains awarded research projects and applications for research grants. It also contains data on components and location of every research project.

d. **Human Resource**. The model adopted for this functional area allows to investigate on the legal-economic careers and wages of the academic personnel. Moreover, it allows to extract information related to the functions, activities, and location of the academic, administrative, and technical personnel.

6 OLAP Layer

The data warehouse supports OLAP queries producing reports for managers and decision makers. In Figure 1, the decision makers are depicted and they are either internal or national decisional agencies. Precisely, we have the following:

a. **Academic Supervisory Staff**. There are two principal Academic Supervisory Staffs: the Academic Senate and the Administration Council. The Academic Senate is the governing body in matter of programming the development of the Athenaeum and coordination of Didactics and Research. It approves the criteria for the distribution of the financings among the Research Structures. Moreover, it determines the criteria for the evaluation of the didactic activities and estimates the effectiveness by analyzing the report produced by the Evaluation Team. This is a partially-elected and independent team, nominated by the University Rector, and it has the function to verify periodically the operating efficiency of didactic structures, research structures and structures for the technical-administrative management. The Administration Council deliberates and supervises the administrative, financial, and economic-patrimonial management of the Athenaeum. In particular, the Council deliberates over the performance of the criteria for the distribution of the financial resources among institutions and the technical and administrative staffs of the University.

b. **Organizational structures**. Faculties are the fundamental structures that organize and coordinate the Didactic activities. In the University, the management of the Research activities is entrusted to the Departments. The Departments are the organizational structures that collect teachers and researchers coming from several Faculties, but joined by the same scientific interests and research methodologies. The Departments collaborate with the Faculties for the realization of the Didactic activities.

c. **Administrative structures**. These structures are the student secretariats and the data elaboration centres, whose tasks are the production of data for the national “Alma Laurea” registry of the graduate students and the editing of documents, statements and other information reports to support the decisional processes.

d. **MIUR**. The national committee for the evaluation of the university system is the MIUR institutional team. Tasks of MIUR are: to establish the general criteria for the evaluation of the activities of the university; to predispose the annual report on the evaluation of the university system; to promote the experimentation, application, and spread of methodologies and evaluation tasks; to determine the nature of the information and data that each athenaeum evaluation team must communicate; to predispose studies and documentation on the state of the university Education, the compliance with the study right, and the access to the university courses of study.

e. **CRUI**. The CRUI is the Association of the Rectors of the Italian Universities. It started up in 1963 as a private association of the Rectors and, in short time, it has acquired a recognized institutional role and a concrete ability to influence the development of the university system through an intense activity of study and experimentation. In particular, CRUI centralizes its own evaluation activity on Didactics and Research areas, and develops and proposes methodologies and evaluation criteria for athenaeums, and degree courses, finalized to improve the quality in the Italian university system.

7 The Didactics Data Mart

This Data Mart has been designed by integrating the logical schemas of two transactional databases: (a) ESSE3, the current database that supports all the didactic curricula, and administrative processes and services to the students; and (b) NOGE the secondary database that stores residual historical data about students enrolled before the ESSE3 introduction.

NOGE and ESSE3 constitute also the repository of data used to feed the Data Mart by the *Extraction-Transformation-Loading* (ETL) process, which is the step in which data are loaded from (huge) source tables into target tables (data marts). There are significant differences between the two databases, and this makes very difficult the process of data integration. The difference regards not only the cardinality of tables but also the data representation.

ESSE3 is a database where the cardinality of tables is of the order of millions of tuples. Errors can be found only in string fields and generally consist of typos. In NOGE, there are tables with thousands of records, but, in spite of this, the database contains very dirty data. Typical errors are nulls and typos, due to particular formats allowed for data and the absence of control during the data entry phase. For example, dates are usually stored as strings *yyyymmdd*, and thus allowing the user to insert partial and/or inconsistent dates with no possible control by the DBMS. Moreover, the most serious problem regards the presence of redundant/duplicate records, often inconsistent among them. For example, the examination of a student on a specific study course might be recorded two times reporting the same date and mark, but different values in the *cum laude* field.

Finally, the two databases adopt different data entry standards. (For example, the tables that store data about the Departments of the Athenaeum could differ significantly in the values naming the structure. It is possible to store “Informatics Department” and “Department of Informatics” indistincly).

All these problems have been faced with data cleaning, that is a sub-process of the ETL process where it is necessary to solve problems at the instance-level relative to actual data contents which are not mageable at the logical schema level. In general, misspellings, typos, abbreviations, and conflicting representation errors can be identified and corrected by a spell checking, based on dictionary lookup that includes synonyms, and by transforming all the data in a standardized format. Duplicate record eliminations require to identify similar records that refer the same real world object, by “fuzzy matching” techniques based on matching rules [13].

Figure 2 reports the detailed logic model of the Didactics data mart. The Didactics data mart contains the following fact tables: enrolment, tax, examination, and degree. All these fact tables have three dimension tables in common: student, degree course and time. These are the basic dimensions, because they represent the minimum of information to express «who-where-when» aggregation levels.

The **enrolment** fact table has five dimension tables. The additional dimensions are: residence,

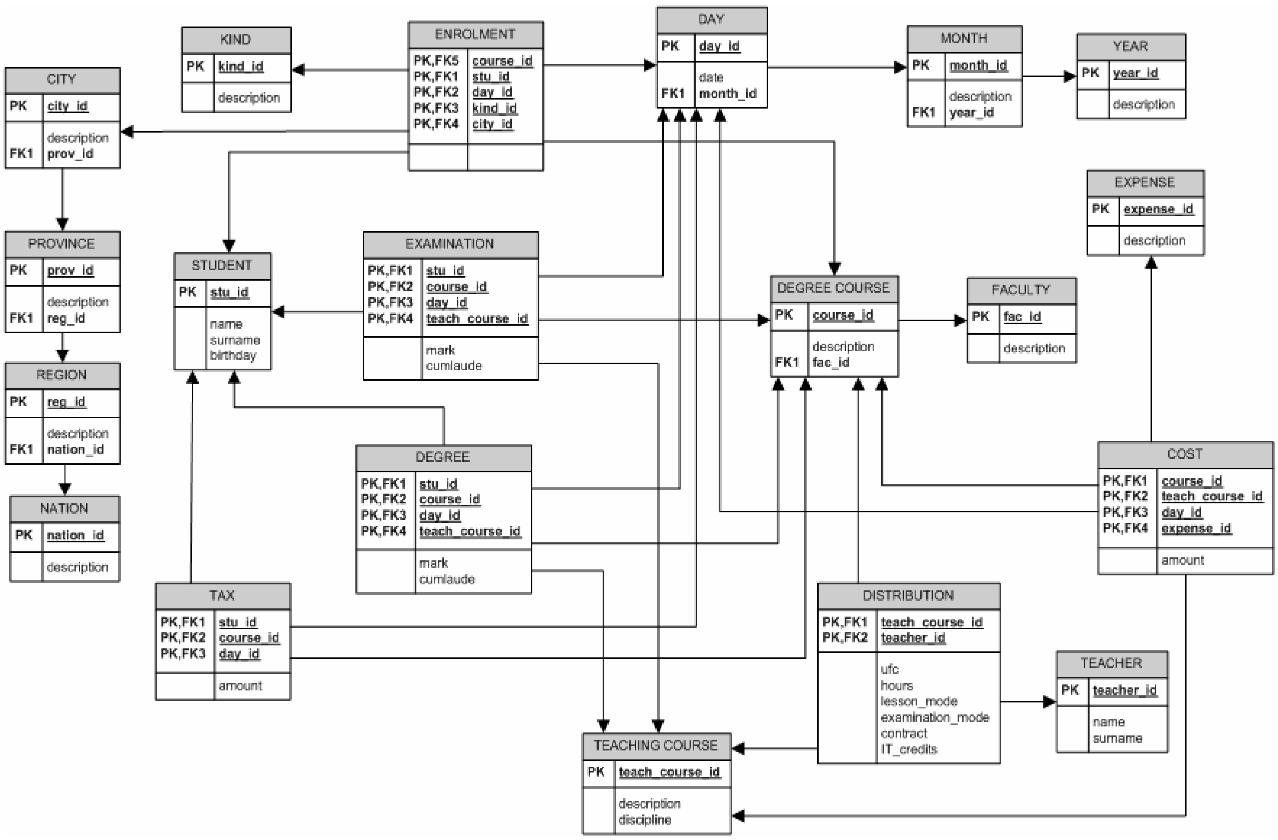

Fig 2. The Didactics data mart.

that allows demographic or geographic aggregation, and kind of enrolment, that allows administrative aggregation. This table has no measures and its function is to store the enrolment to a course of study by the student.

The **tax** fact table has only the student, time and degree course dimensions and it has the amount field as measure. Its function is to control the payment of the taxes by the student.

The **examination** fact table has four dimension tables; the additional dimension is represented by the teaching course, that allows didactic aggregation. It has two fields: the first field is the mark, that represents the fundamental measure; the second is the *cum laude* field, that is a Boolean field.

The **degree** fact table has four dimension tables too; it has the same additional dimension owned by the examination fact table and it has the same measure and fields: mark and *cum laude*.

Moreover, in this data mart there are further fact tables to allow analyses and statistics on the didactic offer of the University.

The **distribution** fact table indicate the list of teaching course for each degree course of study. The information includes the number of teaching hours, the number of university formative credits (UFC), the kind of lesson, the kind of examination, and also the teacher of each teaching course.

The **cost** fact table is relative to the annual costs supported by the University for the management of each teaching course, and totally for each degree course per academic year. It contains also information on the teacher's cost, when the teacher is not enrolled in the University teacher's staff.

In order to obtain aggregate results at a different level of granularity, some dimensions are organized in dimensional hierarchy. In particular, the degree course is a two-level dimensional hierarchy: course and faculty, for allowing aggregate measures (for example, the count of graduate student) at the degree study level or faculty level. The residence is a four-level dimensional hierarchy for aggregate measures at the city, province, region, and nation levels for analyzing data aggregation referring to different geographic contexts. Finally, the time is a three-level dimensional hierarchy including the day, month and year levels, for grouping data respectively by day, month or year. All the other dimensions of the didactic data mart are one-level hierarchies.

8 Example of Academic Applications

Data analyses with OLAP and data mining techniques are used to achieve reports and responses to complex queries. The crucial component of the decisional activity is the summary data section in which a set of statistical indicators regarding the organization data warehouse gives information on the progress of the specific field for which it has been planned. For example, referring to the Didactics data mart, goals of investigation can regard information on the student, forecasts of the formation trend of the Athenaeum, such as:

- Monitoring the incoming and outgoing flows of the students in the University: the count of matriculated students grouped by academic year, the count of enrolled students grouped by academic year and course year (distinguished in regularly enrolled students or outside run students), the count of graduated students by session and degree course, the count of successful examinations and the average mark by academic year and teaching course, the average count of successful examinations by course year.
- Monitoring the didactic workload of the teaching staff: number of hours per teacher and didactic activity, number of presences in commission of profit or degree examination with various roles, number of reported degree theses.
- Monitoring the financial trends of the student taxes.
- Monitoring the needs of a teaching subject matter (Informatics, Mathematics, Foreign Languages, …) in the didactic offer of the University.

We model the decisional making phase as a *Goal Question Indicator Metric* (GQIM) [10]. So, we start from the already defined goals of investigations to produce a set of questions for each of them. A question serves to drive our analysis to real aspects. Answering to a question means getting information and discovering knowledge hidden in the data of the repository. In order to provide the answers to the questions, it is important to define a set of indicators. An indicator represents a way to define how to collect information at a high level of abstraction. Then, every *indicator* must be traduced into a *metric*, by defining an aggregate function on a specific parameter. Finally, the reports calculate the *metric* and show the data to answer to our business questions.

Now, in Table 1, there are reported examples of main *indicators* used for analysis with the respective *metrics* that the analysis and reporting phase produce, with reference to the business goals.

Table 1. Example of GQIM.

Goal	Question	Indicator	Metric
Improve the quality of the Didactics	What Faculties have students with the lowest performance	Mean value of examination marks	Average (examination mark) grouped by Faculty
		Mean value of degree marks	Average (degree mark) grouped by Faculty
Monitoring the incoming flows	Which are the most important Faculties	Number of students belonging to Faculties	Count (stud_id) grouped by Faculty
…	…	…	…

In order to answer to the listed questions, the system can produce traditional statistics as those reported in Tables 2, 3, and 4.

Table 2 shows the results of the Mean value of examination marks indicator. The corresponding metrics is the mean value obtained by the university students in the examinations. The values have been grouped by Faculty and academic year to obtain information about the name of the Faculty whose students pass the examinations with the highest score.

In the last considered year (the 2005 year), it seems that the students of the Arts and Philosophy Faculty have the best mean value. In fact, if we look at the complete series of values, from 2002 to 2005, the students of this Faculty have ever shown a good trend.

It is possible to say quite the same thing about the students of the Educational Science and Foreign Languages and Literature Faculties, although the mean values of the students of others Faculties are not so different from those.

The worst university students seem to be those of Economics of Brindisi city Faculty, which have a very low mean value and even the lowest value in 2003 year. Also the students of Pharmacy Faculty have a low mean

The results shown in Table 2 are confirmed by Table 3, that shows the *Mean value of degree marks* indicator. The metrics calculated is the mean of the degree marks grouped by Faculty from 2002 to 2005. Also in this report, the students of Arts and Philosophy, Foreign Languages and Literature, and

Table 2. Examination average mark grouped by Faculty from 2002 to 2005.

Faculty	2002	2003	2004	2005
Agricultural Sciences	25.30	25.48	25.50	25.54
Agricultural Sciences (Foggia city)	28.00	26.54	26.11	26.85
Arts and Philosophy	27.71	27.76	27.69	27.74
Biotech. Sci.	25.77	25.71	26.93	26.93
Economics	26.30	26.11	26.07	25.74
Economics (Brindisi city)	23.39	23.22	24.74	24.73
Economics (Foggia city)	24.15	25.25		
Economics (Taranto city)	25.95	26.21	25.95	25.76
Edu. Sci.	27.36	27.48	27.49	27.48
Engineering	25.02	25.13	26.69	27.00
Foreign Lang. and Lit.	27.36	27.37	27.10	26.82
Law	25.62	25.72	25.65	25.54
Law (Taranto city)	25.69	25.80	25.57	25.27
Math., Phys. and Nat. Sci.	25.72	25.65	25.93	25.84
Math., Phys. and Nat. Sci. (Taranto city)	26.27	26.45	25.83	25.70
Med. and Surgery	26.46	26.46	26.53	26.63
Med. and Surgery (Foggia city)	25.67	25.48	25.53	25.58
Pharmacy	24.61	24.53	24.49	24.27
Political Sciences	26.51	26.47	26.37	26.14
Veterinary Medicine	25.00	25.09	25.19	25.06

Table 3. Degree average mark grouped by Faculty from 2002 to 2005.

Faculty	2002	2003	2004	2005
Agricultural Sciences	104.89	104.45	104.73	104.02
AgriculturalSciences (Foggia city)	110.00	105.00		108.00
Arts and Philosophy	107.69	107.53	107.83	107.92
Biotech. Sci.			107.70	107.47
Economics	104.81	105.13	105.10	105.32
Economics (Brindisi city)			101.75	101.26
Economics (Foggia city)		92.50		110.00
Economics (Taranto city)	106.30	104.53	104.93	105.07
Edu. Sci.	107.26	107.01	107.57	107.09
Engineering	105.29	96.47	102.57	110.00
Foreign Lang.and Lit.	107.10	106.95	107.50	107.16
Law	101.62	101.76	101.80	101.17
Law (Taranto city)		105.71	103.51	101.78
Math., Phys. and Nat. Sci.	104.11	104.00	104.24	104.28
Math., Phys. and Nat. Sci. (Taranto city)	107.33	107.91	104.82	104.41
Med. and Surgery	106.92	106.12	106.82	105.53
Med. and Surgery (Foggia city)	106.00	103.33		
Pharmacy	100.58	98.61	100.23	99.89
Political Sciences	104.07	103.69	104.39	103.88
Veterinary Medicine	100.64	101.07	101.67	101.26

Educational Science Faculties show a good trend, with quite the same values.

By comparing Table 2 and Table 3, we can surely say that the students of Pharmacy Faculty are the graduates with the lowest performance at all.

Table 4 shows the *Number of students belonging to a Faculty* indicator. The relative metric is for each faculty from 2002 to 2005. It is evident that the humanities Faculties, such as Law, Educational Science, and Economics, produce a large number of graduates every year. In fact, traditionally, a lot of students have always chosen the Law or Economic Faculties, that nowadays are, for social and historical reasons, the most populated Faculties of the University.

In fact, as we see in Figure 3, Educational Science and Law are the Faculties with the highest percentage of enrolled students.

We observe that, as Faculty dimension is a father dimension of Degree Course, the same analysis with a drill down operation, provides a finer-grained view that allows to consider the examination average mark for each Degree Course of all Faculties.

Table 5 shows an example of drill down on *Mean value of degree marks* indicator, shown in Table 3. This new report calculates the mean of the degree marks grouped by Faculty and Degree Course, where the Faculty is Agricultural Sciences.

Table 4. Number of graduate students grouped by Faculty from 2002 to 2005.

Faculty	2002	2003	2004	2005
Agricultural Sciences	271	285	266	350
Agricultural Sciences (Foggia city)	2	2		2
Arts and Philosophy	1272	1229	1299	1327
Biotech. Sci.			54	136
Economics	2381	2333	2334	2514
Economics (Brindisi city)			8	38
Economics (Foggia city)		4		2
Economics (Taranto city)	77	112	109	127
Edu. Sci.	1322	1310	3272	3092
Engineering	14	17	14	2
Foreign Lang. and Lit.	712	600	982	976
Law	3324	3154	3355	3351
Law (Taranto city)		14	136	226
Math., Phys. and Nat. Sci.	1602	1660	1994	1972
Math., Phys. and Nat. Sci. (Taranto city)	48	68	44	44
Med. and Surgery	1134	1006	1414	1476
Med. and Surgery (Foggia city)	4	3		
Pharmacy	496	584	462	476
Political Sciences	932	882	836	908
Veterinary Medicine	174	241	171	178

Complex analyses are accomplished as shown in Informatics credits in various university degree courses.

The aim of this analysis consists of listing all Informatics teaching needs in the University degree courses, showing the UFC credits, teachers, and – if Table 6, that reports the presence of the allowed – equivalent I.T. certifications. Analyzing this report, the Academic Senate obtains indicators on the quality level and teaching efforts (in term of teacher's and teaching costs) about the overall Informatics teaching in the University.

Table 5. Degree average mark grouped by Faculty and Degree Course from 2004 to 2005.

Faculty	Degree Course	Average	
		2004	2005
Agricultural sciences	Technical and Economic Managements of the Rural Land	110.00	104.67
	Medicine of the Plants		110.00
	Animals Productions in Agricultural Systems		95.50
	Plants Productions	106.67	100.67
	Agricultural Sciences	101.56	100.67
	Food Sciences and Technologies		108.00
	Agricultural Sciences and Technologies	104.45	105.45
	Forestry Sciences	100.40	103.40
	Forestry and Environmental Sciences	105.68	104.28
	Phytosanitary Technologies	106.33	99.43
	Technologies Transformations and Quality of the Agriculrtual and Food Products	108.50	105.91
…	…	…	…

9 Conclusions

The paper summarizes the experience in designing and modelling a university data warehouse. Existing facilities and databases affect the chosen data warehouse, that brings them together to support decisional activities leading the whole university environment, including administrators, Faculties and students. The choice to develop a dedicated system is mainly forced by the peculiar information type that defines the basic information in data warehouse widely different from institution to institution.

Future work will provide the extension of the system with a high-performance layer for describing and managing data profiles in the warehouse [14].

This will be done in order to support approximate query processing for OLAP applications that allows fast analytical query execution.

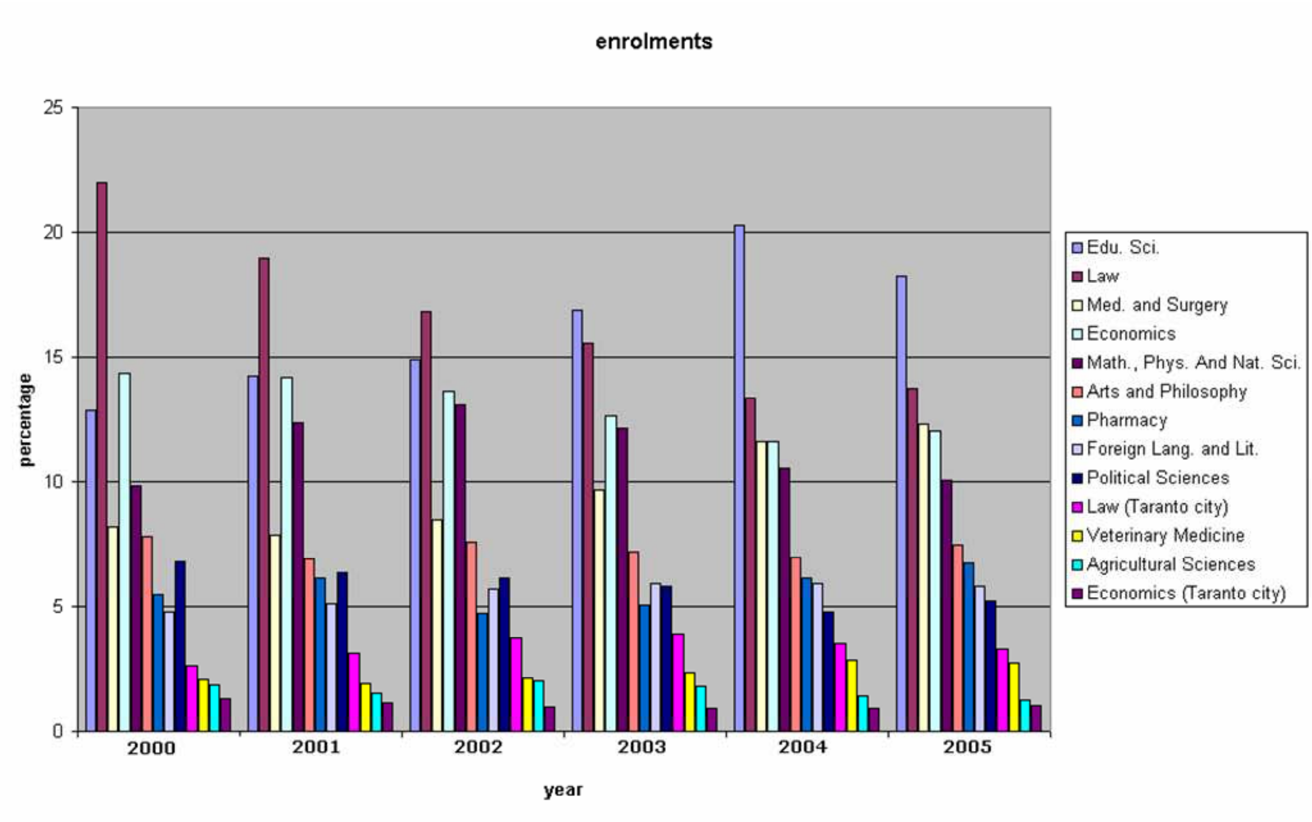

Fig 3. Percentage of university students, grouped by Faculty and Year (2000 - 2005).

Table 6. Informatics Teaching Courses of Degree Curricula.

DEGREE COURSE	TEACHING COURSE	DISCIPLINE SECTOR	LECTURE UFC	LECTURE HOURS	LAB UFC	LABORA-TORY	TEACHER			ALLOWED I.T. CERTIFICATION	
							STRUC-TURED	DISCIPLINE SECTOR	UNDER CONTRACT	ECDL	OTHER
Physics	Informatics Fundamentals	ING-INF/05	6	48	2	base	Y	ING-INF/05			
Physics	Programming Languages	ING-INF/05	1	8	2	specialist	Y	INF/01			
Cultural Heritage	Informatics Applications	INF/01	6			none			y		
Cultural Heritage	Informatics	INF/01	4	32	2	base			y		
Mathema-tics	Informatics	INF/01	7	42	2	base					
Bio-sanitary Science	Informatics	INF/01			3	base				y	MS Certif.
…	…	…	…	…	…	…	…	…	…	…	…

Acknowledgment:
We wish to thank the CSI (Informatics Service Center) of the University of Bari for the valuable support in analysing the system.

References:

[1] K. Curko, M. Varga, and A. Loncar, The support of business intelligence technology in Process and Business Engineering, *International Journal of Computers*, Vol. 2, 2008, pp. 8-14.

[2] W. H. Immon, *Building the data warehouse*, New York: John Wiley & Sons, 1996.

[3] S. Chaundhuri, U. Dayal, and V. Ganti, Database technology for decision support systems, *IEEE Computer*, Vol. 34, 2001, pp. 48-55.

[4] M. Jarke, M. Lenzerini, Y. Vassiliou, and P. Vassiliadis, *Fundamentals of data warehouses*, New York: Springer-Verlag, 2003.

[5] R. Kimball and M. Ross, *The Data warehouse toolkit 2nd edition*, New York: John Wiley & Sons, 2002.

[6] D. Wierschem, J. McMillen, and R. McBroom, What academia can gain from building a data warehouse, *EDUCAUSE Quarterly*, Vol. 26, 2003, pp. 41-46.

[7] G.L. Donhardt and D.M. Keel, The analytical data warehouse: empowering institutional decision makers, *EDUCAUSE Quarterly*, Vol. 24, 2001, pp. 56-58.

[8] M.C. Lin, University data warehouse design issues: case study, in *Proc. of the 2001 American Society for Engineering Education Annual Conference & Exposition*, Albuquerque, 2001, pp. 1-9.

[9] C. Fernandes and M. Whalen, Data warehousing from the web, in *Proc. of the 2004 American Society for Engineering Education Annual Conference & Exposition*, Salt Lake City, 2004, pp. 1-11.

[10] Laila Niedrite, Darja Solodovnikova, Maris Treimanis, and Aivars Niedritis, The development method for process - Oriented Data Warehouse, *WSEAS Transactions on Computer Research*, Vol. 2, 2007, pp 183-190.

[11] CINECA - Consorzio Interuniversitario, Bologna. Available: www.cineca.it.

[12] C. dell'Aquila, F. Di Tria, E. Lefons, and F. Tangorra, Business intelligence systems: a comparative analysis, *WSEAS Transactions on Information Science & Applications*, Vol. 5, 2008, pp. 612-621.

[13] E. Rahm and H. Hai Do, Data cleaning: problems and current approaches, *IEEE Bulletin of the Technical Committee on Data Engineering*, Vol. 23, 2000, pp. 3-13.

[14] C. dell'Aquila, E. Lefons, and F. Tangorra, Decisional portal using approximate query processing, *WSEAS Transactions on Computers*, Vol. 2, 2003, pp. 486-492.