Using Microsoft Analysis Service to analyze graduates' performances and working conditions

Alberto Leone[1], Loris Cancellieri[2], Angelo Guerriero[3] and Andrea Cammelli[4]
Consorzio Interuniversitario AlmaLaurea, Viale Masini 36, 40126 Bologna (BO), Italy
[1]alberto.leone@almalaurea.it [2]loris.cancellieri@almalaurea.it [3]angelo.guerriero@almalaurea.it
[4]andrea.cammelli@almalaurea.it

Keywords
Data Warehouse, Analysis Services, AlmaLaurea, Quality Assessment

1.1. Abstract

Since 1994 several Italian Universities have been gathering invaluable information about their graduates. They founded a Consortium, AlmaLaurea, which had the mission to collect data in a central database, submit surveys about working and studying conditions after graduation. Presently 60 Universities (representing 70% of Italian students) have joined AlmaLaurea Consortium.

Every year AlmaLaurea releases a detailed report about the Italian graduates' profile and their working conditions. The report consists of a qualitative analysis together with a detailed quantitative datasheet. Measurements and statistics can be evaluated from a bird-view perspective aggregating together the records of all the universities to the deeper granularity (e.g. comparing male and female performance and condition for each single course of study). AlmaLaurea represents one of the fundamental sources of information about the Italian University outlook.

AlmaLaurea has been collecting a large amount of information during its activity. Beside the main graduates' database and related survey results, we can find other precious sources: the activity of companies on the CV database, the job offer posting and the job application sent back by graduates. Using all these information, we are able to estimate the preferences companies use to select curricula, the characteristics of the degrees owned by graduates selected from the database, the most attractive degree, the professional figures mainly required by companies, etc

As a consequence you can answer to a lot of questions about the labour market related to University graduates. You can estimate how companies behaves as a function of their dimension, theirs field of activity, their location. The number of questions is virtually infinite.

In order to answer to those questions, we should be able to retrieve and analyse data in a smart way. Therefore we need to:

i) organize all these information in a structured and well documented data warehouse;
ii) introduce the right tools to query and analyse them.

This work is the natural follow-up to the paper presented during the last EUNIS conference in Santiago de Compostela (Leone A. et al, 2009). There we addressed the first issue, introducing a list of all available data sources and starting the building of the data warehouse. It was necessary to face issues of different kind: technical, organizational and even legal, like the respect of the regulation about privacy. Here we present the further work. How the structure of the data warehouse is growing and evolving according to Kimball theories (see Figure 4). Moreover we are experimenting analysis tools, like Microsoft Analysis service, to query and analyze the data.
We will describe the logical structure of the database, the nature and the content of dimensions and fact tables and how these concepts have been mapped into the Microsoft Analysis Services model. To better understand the potentiality of the instrument, we are designing the report structures together with people working on quality assessment and statistical observatories inside the 60 partner Universities.

1.2. Introduction - AlmaLaurea

AlmaLaurea was founded in 1994 on the initiative of the Statistical Observatory of the University of Bologna in order to provide innovative services to graduates, companies and universities. Today AlmaLaurea is a Inter-University Consortium of 60 Italian Universities (the overall number of public and private Italian Universities is about 80) supported by the Ministry of Education, University and Research. Every year about 180.000 new graduates' records (75% of Italian graduates) are added to the database. Presently the database collects 1,540,000 graduates' records and 1,370,000 is the number of CVs available to the employers.

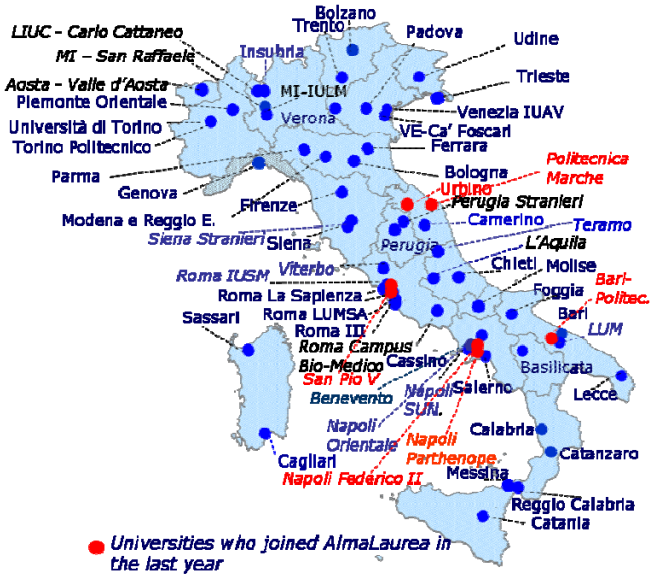


Figure 1. Universities in AlmaLaurea (2009)

AlmaLaurea was set up with the aim of building a National statistical observatory of the University and its relationships with the enterprise world. AlmaLaurea, putting businesses and graduates in contact, became a reference point within the university system for the subjects involved (scholars, operators, etc.) in university education, employment and the development of young people in general.
Every year AlmaLaurea publishes two analyses on Italian graduates:

- **Annual Graduate Profile Report.**
  The Report examines all the graduates of the year, considering their characteristics and performances in the light of a multitude of variables including age at graduation, continuity of studies and attendance, parents' education, social background, study abroad, apprenticeships or internships, foreign languages and IT skills, etc. (AlmaLaurea, 2010-1).
- **Annual Report on the Occupational Conditions of Graduates.**
  The Report provides in-depth information about the employment conditions of young graduates at one, three, and five years from completion of studies, the prospects of the labour market and the relationships between university studies and employment opportunities (AlmaLaurea 2010-2).

2. Overview of available Data Sources

The central data warehouse of AlmaLaurea contains information originating from different data sources and subjects. The main source is the relational database of AlmaLaurea, which manages and helps personnel in the process of verifying, validating and cleaning data, which come from two entities:

- The **universities** taking part in the Consortium which send data concerning the title achieved, the university career, the thesis which has been carried out and previous educational qualifications (including the high school diploma)

- The **graduate** community which, through the filling in of a detailed on-line questionnaire, provide data pertinent to linguistic and computer related expertise, possible working experience during studies and type of employment which is sought. Besides this, they provide an accurate assessment of the university experience which has been carried out

The annual reports are made available, with different levels of details, to universities, to the Ministry of university, to the press and to all the visitors of the AlmaLaurea website.

Over the years the number of reports produced and of variables analyzed has grown considerably, in order to satisfy the demands of very different subjects requiring statistics with different levels of granularity. The reports allow simple, but highly detailed comparisons among universities, faculties, disciplinary areas and, at the finest detail, the specific course of study. Data can be compared also along the time.

After that the aforementioned data have been checked and cleaned (Leone A. & al, 2008) all the collected information is made anonymous and integrated into the warehouse.
A very well identified subset of the available information is used to build a rich and structured curriculum vitae. Graduates have also the opportunity to publish their CVs on-line in the AlmaLaurea website to favour their access to the labour market. In fact, companies can search and browse the online database. Graduates can modify their CVs, adding information about post-graduation training courses and work experiences.

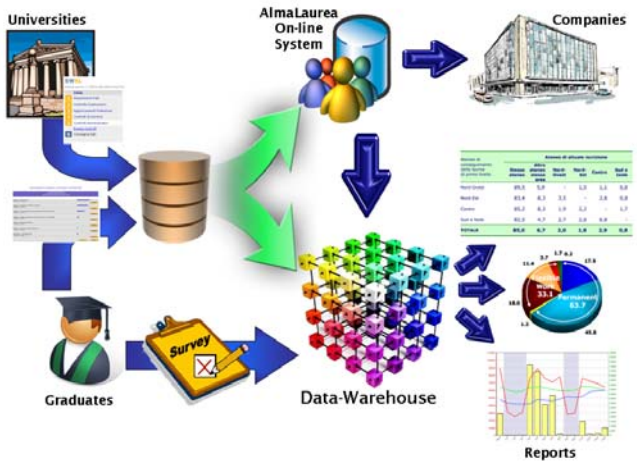


Figure 2. Schematic Data Flow in the AlmaLaurea Data Warehouse

The analysis of the efficiency of the University system is enriched by data provided by various interviews regarding the employment and studying conditions after graduation. The surveys are conducted in a systematic way in the years following the degree completion (Cammelli A., 2008). These interviews are conducted through CAWI (Computer Aided Web Interview) and CATI (Computer Aided Telephone Interview) methodologies.

Also information derived from activities related to the on-line CV done by companies and graduates themselves are conveyed in the DW. We can mention data originating from the distribution of CVs to companies, from the publishing of work offers on the notice-board of the website and from the spontaneous application of graduates to these offers.

To give an idea of the number of the different analyzable combinations one can consider that the cube of the analysis of the graduates of the solar year of 2009 from 52 different universities taking

part in the consortium (about 190.000 graduates) consists of 13 dimensions and, if all significant aggregations were calculated, one would have more than 6.000.000, each of which would count more than 100 calculated members.

The same kind of analysis and report is done yearly to analyze the employment condition of graduates one, three and five years on after graduation. The acquisition of the data necessary to this kind of analysis is done through CAWI and CATI surveys.
Also in this case reports having different levels of granularity – up to the level of the single degree courses are produced. The possibility to supply significant statistics with a very fine level of granularity is guaranteed by the high number of graduates that we try to contact for the survey and also by the fact that the answer rate is always over 70% , with a maximum of 95% for the cohorts that are the most willing to answer.

In addition to these data-marts, which can also be consulted publicly, there are other ones which are currently available only to the internal researchers. As for the reports – which are still being elaborated and validated- they will be available to the universities which will apply for them through an extranet system. Through these reports it will be possible to analyse several factors related to the quality of the human capital produced by universities, such as:

- The sectors which are mostly required by companies in a given period
- The possibility to be selected on the basis of the age/ age at graduation
- The importance of factors such as foreign languages skills and diplomas certifying them.
- The characteristics of the CVs which are the most read by companies and their dependence from factors such as: the company economical sector, the main office and the dimension of the company.

3. Data collection process

The data warehouse collects data from several and heterogeneous sources. Some data are collected by structured questionnaires filled online by the users themselves (CAWI- Computer Aided Web Interview) or by an interviewer (CATI – Computer Aided Telephone Interview). Other data are collected by relational databases, where the activities of the different users are recorded: search parameters used by the employers when browsing and searching the CV database, the CVs that have been selected and downloaded by employers. Other data are collected directly by the Universities, such as the characteristics of all the degrees.

Type	Quality/Reliability	Transformation	Source	Formats
Interviews	Very high	None	Local	Relational data
Administrative Data	High	Cleaning, Trans-coding	Remote, heterogeneous	XML
Enterprise information system	Medium	Cleaning, trans-coding	Local/Remote	Relational data, XML
Log data	low	Filtering	Local/Remote	Relational data, text files, XML
Mailing Data	Medium	Cleaning, Trans-coding	Remote	Text/cvs data files

Table 1. Data source categories

This is the list of the different types of data, grouped by type:

1. Interviews.
2. Administrative data. Collected from Universities. Extracted from several and different University Information systems
3. Enterprise information system. Data extracted from relational databases.
4. Log data. Other data about user activities recorded into system and application logs.

Each category of data is characterized by different quality and reliability, that ETL procedures have to take into account. Moreover we have to deal with different data formats and transferring procedures.

4. Data Warehouse implementation

The data warehouse implementation is still in progress. We are building the inventory of all the potentially useful data sources and related metadata. The acquisition of a new data source is performed following different steps. At the end of the process a semi-automated procedure allow for periodical import and update of new data.

Source	Type	Description	Update Frequency
Employment Conditions	interview	CAWI and CATI interviews about graduate employment and studying conditions	yearly
Graduates' profile	interview	CAWI interviews about the university experience	Yearly
Graduates' administrative data	Administrative Data		quarterly
Purchased CVs	Enterprise information system		Daily
Job Offers published	Enterprise information system	Job Offers published by Employers in our bulletin board	Daily
CV's sent by graduates to Job offers	Enterprise Information System	CV's sent by graduates in response to a Job Offers	Daily
Placement Dpt. Activities	Enterprise Information System	Employers contacts, sold CVs and subscriptions, turnover, etc	Daily
Graduates Activities	Log Data	Access and Update of CVs, Search of postgraduate courses	Daily
Mass Mailing Activities	Mailing Data	Sent Messages, Number of recipients, Number of reads, of clicks etc	Daily

Table 2. Data source list (partial)

4.1 ETL modules

Extract, transform and load is the first process that you need to execute in order to use data in a data-warehouse environment. This process consists, mainly, of three phases:

- Extracting data from external data sources (file, relational databases, web-services etc)
- Transforming them to fit operational needs (including quality levels)
- Loading them into the final target (data-warehouse)

As shown in the table above, we deal with data sources that are so different in terms of quality of data that we need to project different ETL procedures for each kind of data-sources. The standard process of an ETL process is:

1. Cycle initiation (preparation of staging environment)
2. Extract data from original sources into bulk load tables
3. Validate the data and make some preliminary upgrade on data
4. Load Into the Staging Tables
5. Transform (clean, apply business rules, check for data integrity, create aggregates or disaggregates)
6. Audit reports on compliance with business rules or, in case of failure, on errors
7. Publish the data to the final tables
8. Archive bulk loaded data
9. Clean up the staging environment

Example 1: Loading of Web Sites Log data

To manage our website we use four different web server load balanced by twos, one couple uses Apache and hosts PHP and PERL applications while the second couple uses IIS and hosts ASP.NET applications. Every machine has its own web log and, consequently, we need to join all the files together in order to analyze them. After that we have to bulk load the information cleaning the dirty ones, to classify the new pages and to group them into visits. Finally the data are loaded into the final relational tables designed using the star-schema.

1. Truncate of the Staging Environment
2. Data are extracted from log files and loaded into bulk load tables

3. Data are splitted into fields. Incomplete lines are deleted. Fields that exist only in one web-log format are deleted from bulk tables. All records that are not relative to page request (i.e. JavaScript files, css file, images, etc) are deleted. The same thing happens to records that have not a 200 (OK) status code.
4. Data are loaded into the staging tables
5. Some transformations are made on data in order to uniform their format (i.e. date/time format, URI format, etc). After that, data are sorted in order to isolate unique visits on website and staging dimensional tables are populated. During this step data are coded in order to populate the fact tables and the dimensional ones. New pages are classified following a set of rules.
6. Using a web-application, staff users could verify the previous activities and modify page, host or user agent classification.
7. Data are loaded into the final dimensional and fact tables
8. The data contained into the staging tables are archived
9. The staging environment is cleaned

All these steps are made using a Sql Server Integration Services package.

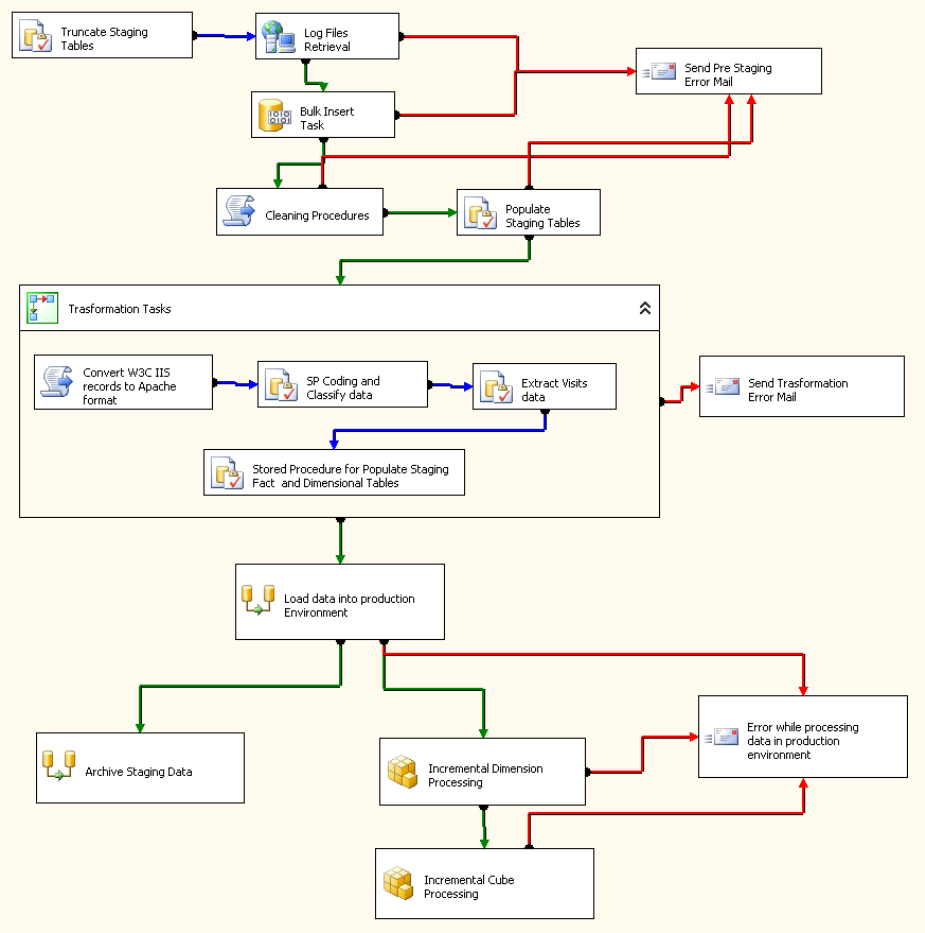


Figure 3. ETL diagram for the web traffic data mart

5. Core data warehouse structure

Figure 4 represents a portion of the data warehouse showing different fact tables, their dimensions and relationships. The main goal of this project is the integration of all the data marts around the main fact/dimension table: the graduates table. Each activity around the AlmaLaurea systems can be referred to a set of graduate records (table Graduates): interview on Employment Conditions (table Employment Conditions), CVs search and download by employers (table Purchased CV).

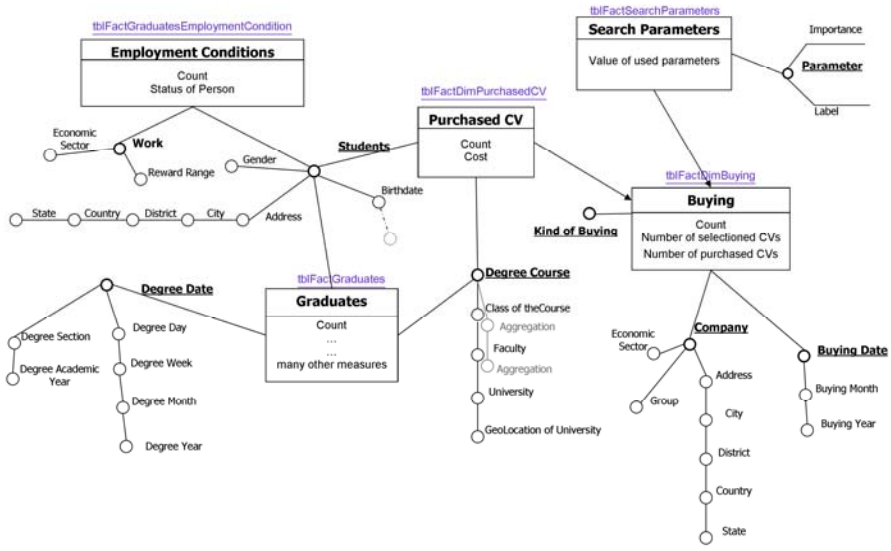


Figure 4. Schematic representation of Fact Tables, Dimensions and Hierarchies

5.1 Metadata

An essential component of a data warehouse is the metadata and tools to manage and retrieve them. Ralph Kimball describes metadata as the DNA of the data warehouse as metadata defines the elements of the data warehouse and how they work together. According to Kimball (KRT2008), metadata can be divided into 2 similar categories: Technical metadata and Business metadata. Technical metadata correspond to internal metadata, business metadata to external metadata.
Business metadata stores business definitions of the data; it contains high-level definitions of all fields present in the data warehouse, information about cubes, aggregates, datamarts and so on.
Business metadata is mainly addressed to and used by the data warehouse users, report authors (for ad-hoc querying), cubes creators, data managers, testers, analysts.
Typically, the following information needs to be provided to describe business metadata:

- DW Table Name
- DW Column Name
- Business Name - short and descriptive header information
- Definition - extended description with brief overview of the business rules for the field
- Field Type - a flag may indicate whether a given field stores the key or a discrete value, whether is active or not, or what data type is it. The content of that field (or fields) may vary upon business needs.

Technical metadata is a representation of the ETL process. It stores data mapping and transformations from source systems to the data warehouse and is mostly used by data-warehouse developers, specialists and ETL modelers. The most important pieces of information which you must provide to describe technical metadata are source and target databases, source and target tables, source and target columns and data transformations.

In our system we store all the business metadata into a meta-database. The meta-database contains structured information about each table, fields, and related dimensions.
Metadata also collect management information and the history of all the updates to the data stored into the warehouse.

Information about the ETL process, technical metadata, is not programmatically stored yet. It could be pulled out analyzing the structure of the SISS Package

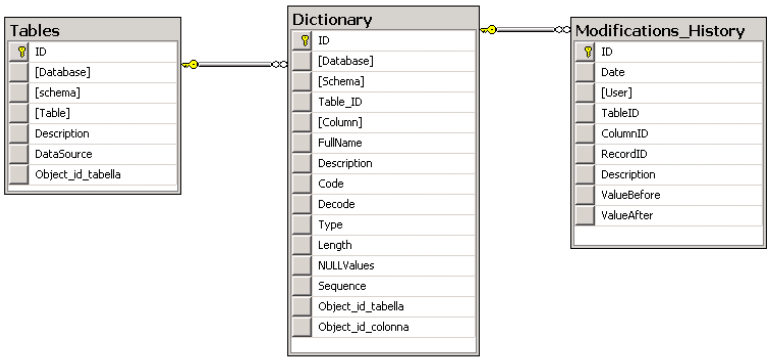


Figure 5. Metadata core diagram

6. Analysis and Reporting tools

Analysis tools permit to make inquiry on the aggregate data (cubes) built by the system. Typically the analysis features of a system lean on OLAP (On-Line Analytical Processing) technology. OLAP is essentially an approach to decision processes that focus on information's dimensional analysis and helps to sole no-structured problems. It uses aggregate data and dimensions because OLAP is oriented to business users and business is made of dimensions rather that of tables. For this reason every business user is able to use an OLAP tool after understanding the concepts of dimensions and hierarchy.
Our solution exploits Microsoft Analysis Services 2005 as OLAP Engine and Microsoft Reporting Services 2008 as tools to explore data. In order to build and deploy projects we use Microsoft Business Intelligence Studio a product that runs into Microsoft Visual Studio environment.

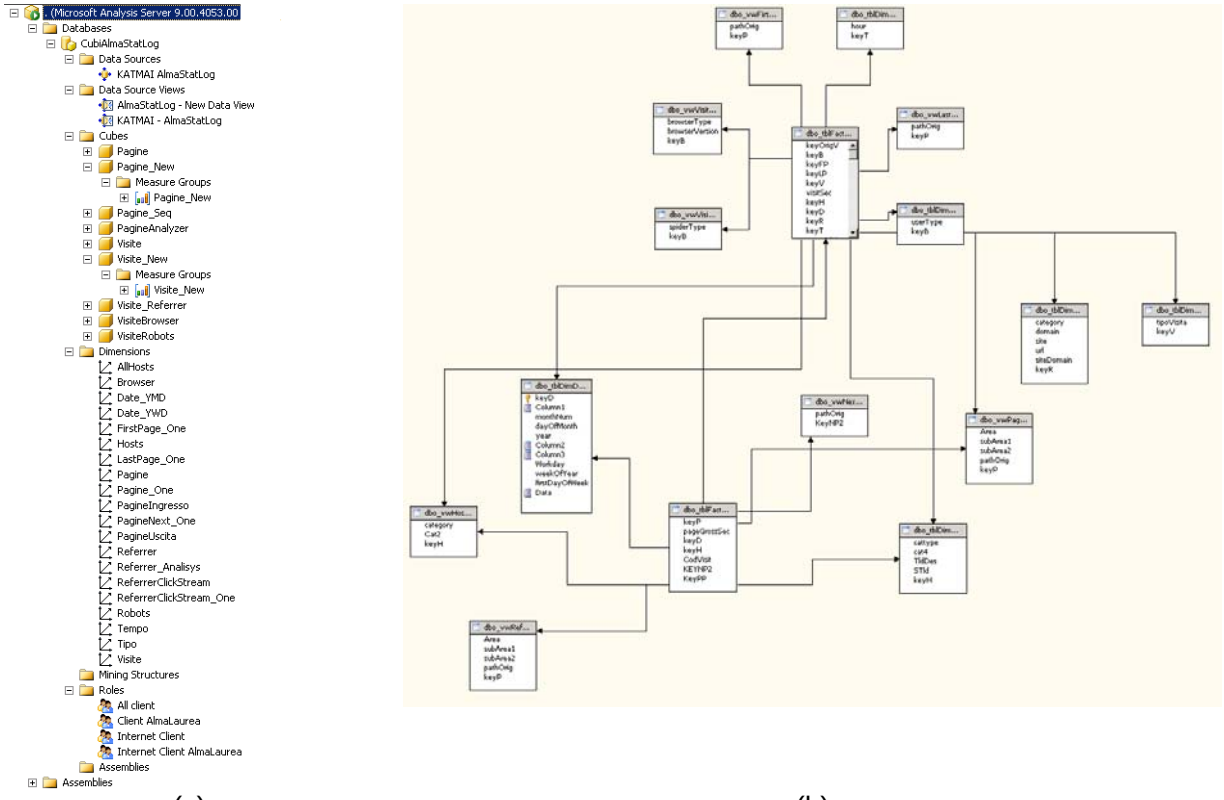
First of all you have to build a new "data view" that is a representation of relational tables involved into the project together with relations, primary and foreign keys between tables. Starting from the data view you could easily build dimensions and hierarchies inside them. You can also define different translations for attribute's names: in that way data could be explored in different languages. After creating the dimensions, you can create the cubes. The steps that you have to follow to create a cube into Microsoft Business Intelligence Studio are:

- Select the "data view" involved in the cube
- Select the fact table and the OLAP dimensions. The attributes of the fact tables must be foreign keys toward the dimensional tables connected to selected OLAP
- Create the measures that you want to calculate, the measures group and any calculate attribute
- Create Cube partitions. Partitions enable you to distribute a measure group into discrete segments on a single server or across multiple servers, and to optimize storage and query performance.
- Choose storage settings. Depending on the version of Analysis Service that you are using you could chose between MOLAP (Pure, Scheduled, Automatic, Medium Latency and Low Latency), HOLAP and ROLAP. The main difference between them is
    a. MOLAP (Multidimensional OLAP) data is stored in a multidimensional cube. The storage is not in the relational database, but in proprietary formats. Advantages: it could perform complex calculations and exploring data (slicing and dicing operations) have excellent performance
    b. ROLAP (Relational OLAP) data remains stored into the original relational database and manipulated to give the appearance of traditional OLAP's slicing and dicing functionality. Every operation relies on generating SQL statements to query the relational database. Advantages: you could handle huge size of data without using other disk space. Disadvantages: responsiveness could be poor (slow query performance) and you cannot perform complex calculations because SQL statements do not fit all needs

c. HOLAP (Hybrid OLAP) attempt to combine the advantages of MOLAP and ROLAP. For summary-type information, HOLAP leverages cube technology for faster performance. When detail information is needed, HOLAP can "drill through" from the cube into the underlying relational data.

- Define any data mining model
- Define any measures translations
- Define security policy (Roles)
- Process the cube in order to test it

When you have finished building all the dimensions and cubes you could deploy the project to the Analysis Service Engine.



(a) (b)

Figure 6. Exploring Analysis Services Objects (a). DataView for Web Log Data Cube (b)

7. Sample one: *delay in degree completion times study*

The example report of Figure 7 presents the *delay in degree completion times study* (*delay_index*) for different disciplinary areas. This index is given by the formula:

$$delay\_index = (duration - legal\_duration) / legal\ duration$$

where *duration* is the duration of studies and *legal_duration* is the prescribed duration defined for that particular course of study. We must remind that in Italy the real duration of a course of studies is often higher than the prescribed duration. For example, in 2008, just 40% of the student graduated within prescribed degree completion time.

The example shows how this figure changes with respect to the field of study and parents' educational qualification. The report has been improved by associating different colours to different value ranges.

Profilo - Indice di Ritardo

Gruppo disciplinare - Titoli dei genitori

Gruppo Statistico	Titoli dei genitori													
	Solo uno dei due con diploma superiore		Solo uno dei due con laurea		Titoli inferiori o nessun titolo		Tutti e due con diploma superiore		Tutti e due con Laurea		Undefined		Totale	
agrario	0.40	708	0.41	482	0.53	1034	0.40	699	0.31	278	0.72	317	0.46	3518
architettura	0.47	1772	0.42	1524	0.61	2086	0.45	2019	0.36	1037	0.81	865	0.51	9303
chimico-farmaceutico	0.35	891	0.39	841	0.49	1267	0.34	1047	0.32	545	0.55	375	0.40	4966
economico-statistico	0.41	4918	0.38	3342	0.54	6396	0.38	5267	0.30	1926	0.79	2052	0.46	23901
educazione fisica	0.33	531	0.31	292	0.33	556	0.30	581	0.33	101	0.30	483	0.32	2544
geo-biologico	0.35	1755	0.32	1289	0.48	1977	0.33	2017	0.30	785	0.53	976	0.39	8799
giuridico	0.67	2233	0.56	2439	0.86	3106	0.63	2827	0.41	1713	1.04	3110	0.73	15428
ingegneria	0.45	4086	0.40	3615	0.60	4382	0.43	4929	0.34	2387	0.74	960	0.47	20359
insegnamento	0.46	1992	0.44	765	0.53	3602	0.42	1510	0.49	250	0.54	1763	0.49	9882
letterario	0.52	3193	0.47	2853	0.72	3796	0.48	3887	0.39	1821	0.81	2737	0.57	18287
linguistico	0.40	2205	0.38	1315	0.59	2556	0.38	2311	0.32	739	0.80	1058	0.48	10184
medico	0.16	3467	0.21	2278	0.18	5617	0.18	3064	0.15	1594	0.27	3258	0.19	19278
politico-sociale	0.38	4872	0.38	3421	0.43	6824	0.36	5435	0.33	1901	0.50	2951	0.40	25404
psicologico	0.34	2129	0.29	1423	0.43	2711	0.30	2330	0.27	826	0.46	586	0.35	10005
scientifico	0.46	1049	0.44	870	0.62	1243	0.40	1206	0.27	631	0.63	502	0.47	5501
Totale	0.41	35801	0.39	26749	0.52	47153	0.39	39129	0.32	16534	0.65	21993	0.45	187359

Figure 7. course duration delay index as a function of field of study and parents' educational qualification (2008 graduates, all the Universities)

Analysing the report, we can immediately see that the medical sector has better performances (lower values for this index), and the result can be very easily explained reminding that the courses of study in this area require an admission examination. In Italy most of the courses of study are open and do not require any examination.
We can also observe the influence of the parent's educational qualification: better results for higher qualification. The index grows from a minimum of 0.32 when both parents are graduated (a delay of 1 year in a course that have a prescribed duration of 3 years) to 0.52 when parents have lower educational qualification or no educational qualification.

Reporting services automatically define a panel where the user can select the subset of data for the report. The panel shows all the variables that are defined as parameters in the queries instanced in the report. For example, we can analyze the same report limiting data to the graduates of the University of Bologna (see Figure 8). The qualitative behaviour remains the same, but the overall delay index is lower than the national one.

Profilo - Indice di Ritardo

Gruppo disciplinare - Titoli dei genitori

Gruppo Statistico	Titoli dei genitori													
	Solo uno dei due con diploma superiore		Solo uno dei due con laurea		Titoli inferiori o nessun titolo		Tutti e due con diploma superiore		Tutti e due con Laurea		Undefined		Totale	
agrario	0.21	94	0.26	67	0.53	87	0.37	90	0.32	49	0.85	32	0.38	419
architettura	0.31	58	0.38	41	0.38	67	0.30	58	0.41	34	0.63	8	0.36	266
chimico-farmaceutico	0.28	63	0.33	102	0.47	95	0.31	81	0.35	68	0.43	19	0.36	428
economico-statistico	0.35	454	0.31	382	0.39	500	0.30	489	0.22	221	0.85	155	0.36	2201
educazione fisica	0.37	47	0.24	32	0.38	47	0.31	52	0.08	12	0.55	39	0.36	229
geo-biologico	0.22	119	0.25	131	0.31	96	0.22	165	0.23	74	0.43	54	0.26	639
giuridico	0.66	163	0.48	225	0.80	173	0.59	231	0.34	177	0.98	161	0.63	1130
ingegneria	0.39	372	0.35	370	0.51	364	0.40	459	0.34	287	0.97	31	0.41	1883
insegnamento	0.30	199	0.49	102	0.38	307	0.31	174	0.51	40	0.54	90	0.38	912
letterario	0.46	424	0.40	459	0.56	394	0.40	492	0.40	293	0.75	627	0.51	2689
linguistico	0.31	169	0.28	125	0.50	163	0.30	177	0.27	89	0.76	54	0.37	777
medico	0.17	195	0.16	168	0.20	242	0.17	186	0.17	136	0.24	112	0.18	1039
politico-sociale	0.30	442	0.26	422	0.47	429	0.30	511	0.26	294	0.51	170	0.34	2268
psicologico	0.12	116	0.19	73	0.28	117	0.15	102	0.13	36	0.66	8	0.19	452
scientifico	0.35	106	0.46	113	0.53	132	0.29	131	0.33	80	0.62	28	0.41	590
Totale	0.34	3021	0.33	2812	0.45	3213	0.33	3398	0.30	1890	0.69	1588	0.39	15922

Figure 8. course duration delay index as a function of field of study and parents' educational qualification (2008 graduates, Bologna University)

8. Sample two: *relationship between field of study and company type*

Figure 9 shows the number of CVs (2008 graduates, University of Bologna) downloaded by companies during 2008. Data are presented by field of study on the rows and company type on the columns. Rows can be drilled down from Faculty level to course of study.
This report is generated combining the fact table of the downloaded CVs and the fact table of graduation (that has the role of a dimension table in this report).

Numero CV Acquistati 5/3/2010 1:25:55 PM

Facoltà	Classe di Laurea	Corso di Laurea	Tipo Azienda: Assicurazioni	Associazioni	Azienda Sconosciuta o di Sistema	Aziende	Banche/Finanziarie	Enti	Enti di formazione	Istituto Superiore	Liberi professionisti	Ricerca personale	Università	Total
⊞ Agraria		Totale	14	2	20	192		9	869			70	263	1439
⊞ Architettura		Totale			14	96		35	86					231
⊞ Chimica industriale		Totale	1	1		114		15	5			3		139
⊞ Conservazione dei beni culturali		Totale	12	2		12	10		74			32	38	180
⊟ Economia - sede di Bologna	⊞ finanza (19/S)	Totale	10	19	4	73	84	6	202			166	31	595
	⊞ scienze dell'economia (64/S)	Totale	29	3	7	189	167	5	626		1	381	69	1477
	⊟ scienze dell'economia e della gestione aziendale (17)	economia aziendale	33 1.07 %	8 0.26 %	4 0.13 %	139 4.51 %	259 8.40 %	24 0.78 %	1179 38.24 %	6 0.19 %	0.00 %	1379 44.73 %	52 1.69 %	3083 100.00 %
		economia e commercio - BOLOGNA	57 15.57 %	2 0.55 %	0.00 %	13 3.55 %	32 8.74 %	1 0.27 %	94 25.68 %	0.00 %	0.00 %	167 45.63 %	0.00 %	366 100.00 %
		economia e gestione delle imprese	3 0.84 %	1 0.28 %	1 0.28 %	24 6.72 %	67 18.77 %	0.00 %	198 55.46 %	0.00 %	0.00 %	47 13.17 %	16 4.48 %	357 100.00 %
		economia e management dei servizi turist ...	11 20.37 %	0.00 %	0.00 %	1 1.85 %	3 5.56 %	0.00 %	28 51.85 %	2 3.70 %	0.00 %	7 12.96 %	2 3.70 %	54 100.00 %
		economia e marketing	42 2.67 %	1 0.06 %	4 0.25 %	99 6.29 %	134 8.52 %	12 0.76 %	888 56.45 %	1 0.06 %	0.00 %	361 22.95 %	31 1.97 %	1573 100.00 %
		economia e professione	13 1.37 %	3 0.32 %	1 0.11 %	61 6.41 %	83 8.73 %	5 0.53 %	402 42.27 %	2 0.21 %	0.00 %	366 38.49 %	15 1.58 %	951 100.00 %
		Totale	159	15	10	337	578	42	2789	11		2327	116	6384
	⊞ scienze economiche (28)	Totale	39	1	7	91	156	21	916			655	56	1942
	⊞ scienze economiche per l'ambiente e la cultura (83/S)	Totale	12	1		25	24	1	263			48	5	379

Figure 9. Companies activity on the CV database by field of study and company type.

9. Technical details

The project is still in progress. While the implementation of the data warehouse and ETL procedures is still in progress, the reporting layer is still under analysis.
Microsoft analysis and reporting tools are a hypothesis that we are still evaluating. At the time of writing of this paper we are still working on data mart design, while the reporting layer is running on development machines with limited resources. We are planning to deploy a first test architecture by the end of year 2010, with a basic set of data mart and access allowed only to AlmaLaurea University members.

10. Conclusions

We presented a case study regarding the use of Microsoft Analysis Services and Microsoft Reporting Services to analyse and present the data contained in the AlmaLaurea Data Warehouse. The data warehouse implementation is still in progress and we are trying different software solutions to build the presentation layer.
Microsoft tools promise to be the best choice combining simplicity, flexibility and compatibility with the technical infrastructure and the application environment of AlmaLaurea.

11. References

AlmaLaurea (2010-1*). Graduate Profile - 2010 Report, Consorzio Interuniversitario AlmaLaurea*, Bologna (see also previous works 1994 – 2009), http://www.almalaurea.it/eng/universita/profilo/

AlmaLaurea (2010-2). *Employment Conditions of Graduates – 2010 Report, Consorzio Interuniversitario AlmaLaurea,* Bologna (see also previous reports 1996-2009), http://www.almalaurea.it/eng/universita/occupazione/

Cammelli A., di Francia A. & Sgarzi M. (2008) *AlmaLaurea Surveys*, EUNIS 2008 Conference Proceedings.

Leone A., Guerriero A., Cancellieri L. & Di Sario P. (2008) *Merging graduates data from remote and non-homogeneous information systems* EUNIS 2008 Conference Proceedings.

Leone A., Cancellieri L., Guerriero A. & Cammelli A., (2009) *The impact of the Bologna Process on the labour market: a national data warehouse regarding universities and graduates* EUNIS 2009 Conference Proceedings

Kimball R. & Ross M., (2002) *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modelling,* 2nd Edition. John Wiley & Sons

Kimball R., Ross M., Thornthwaite W., Mundy J. & Becker B., (2008) *The Data Warehouse Lifecycle Toolkit, 2nd Edition: Practical Techniques for Building Data Warehouse and Business Intelligence Systems* John Wiley & Sons

Kimball R., Thornthwaite W. & Mundy J., (2006) The Microsoft Data Warehouse Toolkit: With SQL Server 2005 and the Microsoft Business Intelligence Toolset John Wiley & Sons