

Business Intelligence as a HRIS for Absenteeism

Alysson Bolognesi Prado, Carmen E. Feitosa de Freitas and Thiago Ricardo Sbrici

Unicamp – State University of Campinas
Rua da Reitoria, s/n – DGRH
Campinas – Sao Paulo – Brazil

Abstract. This paper is based on an experience using Business Intelligence systems for workers absenteeism analysis on a large organization. The related concepts are explained, an actual application case is presented, and finally we discuss some drawbacks and advantages of this approach..

1 Introduction

Modern organizations know the most valuable asset they have are the people who works for them. Therefore, the Human Resources Department must be empowered with tools that best help to hire the proper workers, to reward adequately, to train when necessary, to keep personnel's health and to be benefited of individual interests towards the business goals.

Business Intelligence software were proposed to ease the task of analyzing huge amount of data on the organization, mainly helping taking decisions on market strategies, foreseeing estimated costs and profits, and other information related to customers, suppliers and production.

This paper resulted from the knowledge gathered during a project carried out by the authors in order to model a system to manage the absence of employees from work. We study the concept of Absenteeism and Business Intelligence on Section 2 and 3, respectively. On Section 4 we present an implementation of a Business Intelligence system that address Absenteeism, and Section 5 and 6 finish with result analysis, the drawbacks and advantages of this approach, and some planning for future work.

2 Absenteeism

According to Chiavenato [1], Absenteeism is an expression used to designate worker's faults or absences on the enterprise.

Quick and Lapertosa [2] suggest the following classification, regarding the cause of absence:

- Work sickness Absenteeism: originated by professional diseases or work accidents;

- Health Absenteeism: produced by diseases other than those related to work;
- Legal Absenteeism: when the absences are predicted and accepted by laws;
- Compulsory Absenteeism: situations beyond the worker's will, for instance, any penalty or arrest.
- Free Will Absenteeism: generated by personal affairs and situations without legal subsidy.

Couto [3] emphasizes that is very important to know how to measure and understand the absenteeism rates on a corporation, based not on a single point of view, but instead involving areas as:

- Operational Management: focus on frequency control to detect and avoid problems with team productivity, exceeding hours on duty and overload of tasks.
- Human Resources: build projections to foresee the future needs of working force.
- Occupational Safety and Health: aims to detect and prevent diseases produced by work situation, resulting on programs to deal with hazards and enhance life quality.

2.1 Absenteeism Rates

Based on these approaches, authors define several different absenteeism rate formulas, with slightly variations on the input data, but all having the general rationale of quantifying the absence from work force related to the work force that was expected or desired to be available.

Couto [3] defines his rates based on the average number of employees working and the **number of days** that are important for each analysis: number of days faulted, number of days on medical licenses, and so forth.

The Sipageh Project [4] uses the **number of hours** in unpredicted non-working situations, divided by the contractual number of hours, to obtain a synthetic absenteeism index.

The Bradford factor [5] relies on **the square of** the number of contiguous occasions of absence, multiplied by the number of absences, during the year, to provide a method to weight the short, frequent and unplanned absences, which they consider more harmful to the organization.

Couto [3] highlights the importance of choosing a single standard of absenteeism calculation, in order to make it comparable in terms of time evolution and distribution in the company geography. This convergence can be obtained by a single system that provides the data to the whole organization, however respecting particular needs, by tools as Business Intelligence software described below.

3 Business Intelligence

The concept of Business Intelligence (BI) refers to the abilities of the corporations to retrieve information related to their operation processes and area of activity, in a flexible and dynamic way, allowing the analysis, detailing and understanding their

work and providing means for decision support. The term has been popularized since the late 1980's by Howard Dresner and the Gartner Group [6].

The data managed by Business Intelligence systems have certain specific characteristics, reflecting in the way they are gathered, stored and retrieved, which will be briefly explained in the next sessions.

3.1 Extract, Transform and Load

The process of obtaining and modifying the data for feeding a Business Intelligence database is called ETL, in respect to the three steps it involves: Extract, Transform and Load.

In the Extract step, data are typically queried from other systems of the company, the so-called OLTP – On Line Transactional Processing – that supports the day-by-day organization operations. Spreadsheets and plain text files can also be used as data sources for Extraction.

In the Transform step, the data are handled aiming to fit in the view the users of the decision support system have about the process and facts they represent. This means unity conversions, codes standardization and data filtering, categorization and so forth.

In the Load step the data produced by the prior steps are stored in a special database structure called Data Warehouse, which is described as follows.

3.2 Data Warehousing and Data Mart

The Data Warehouse (DW) is a large data repository [7], obtained from all the relevant sections of the organization. “The Data Warehouse contains the raw material for management’s decision support system”. [8]

When the Data Warehouse is updated from ETL, no data is deleted or overwritten. Instead, the data are accumulated, constructing the history of the data involved on the company’s operations.

The data structure of a DW often does not follow the common database systems techniques, that uses normalization to ensure data integrity and less storage space. Instead, the data are de-normalized and arranged in such a way that helps to query for reports and analysis.

From the organization’s Data Warehouse, the information related to each area of interest is separated, forming Data Marts. Each Data Mart (DM) is a subset of the whole DW focused on a special subject of the decision support systems’ users.

Data on a DM is arranged according to the **star schema**: each relevant data from the organization, known as **fact**, is stored in a structure called **cube**, which can be viewed or decomposed according to pre-defined categorization of the data, known as **dimensions** of the cube. The dimensions are typically some meaningful attributes of the fact, such as time, customer type, products and geography of the organization.

The dimensions can have several hierarchies for classifying the facts according to many points of view, each hierarchy being composed by different levels of granularity.

3.3 On Line Analytical Processing

Retrieving the data from a DM requires specialized software that can deal with cubes and dimensions. Besides, it must be able to provide a single entry point for the managers of the organization to search for information needed to their decision making process. This role is fulfilled by the On Line Analytical Processing systems (OLAP).

An OLAP system provides means for accessing a Data Mart, listing its cubes and related dimensions, constructing reports, independently of the particular data it contains. This way, the user does not need to learn how to operate different systems, and also the programmers do not need to develop new programs and reports for each subject addressed by the Data Marts.

Through an OLAP, a user can build reports and charts that summarize the facts retrieved from the DM, according to the selected dimensions. This way, it is easy to create cross-tabulation by dragging and dropping dimensions on rows or columns.

It's also easy to change the granularity chosen by a certain dimension, in a process called drill-up or drill-down, when the user navigates through the hierarchy levels of a dimension.

For each of these modifications on table structure, the data on the cells of the cross-table are re-summarized on-line, providing immediate answers for the user and helping the decision-making.

4 Implementation

Aiming to identify the absenteeism rate in an organization, we modeled a Data Mart able to supply strategic information designed to facilitate the managers and HR staff to make decisions referring to workers' attendance in the company.

This large organization employs around 10,000 workers, in more than 1,000 departments and offices, some shifts, different number of working hours per day, and one hundred functions.

The project was scheduled to deliver a first implementation that fulfills the requirements of the Operational Management and Human Resources areas, as described on Section 2. Based on the results of its use and feedbacks from users, new versions of the Data Mart must be provided, adding new cubes and dimensions to the DM. This first implementation, regarding only the working force availability and predictability, will be described below.

4.1 Star Schema Suggestive

From an OLTP system used in the company to store and process data related to the daily attendance of the employees, we define a star schema that uses this information, on a consolidated and complete way. The information structure is modeled in agreement with the dimensions that were defined in the project.

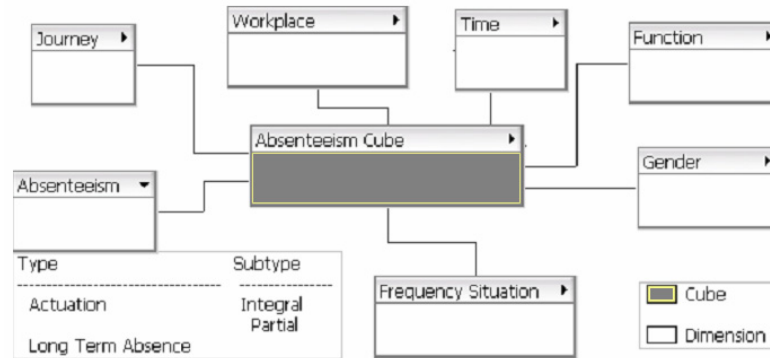


Fig. 1. Star Schema for Absenteeism Data Mart.

Cube. In the beginning of the project we considered the use of the number of days of absence [3] as the fact for the cube. But due to characteristics of the organization, such as the variety of shift durations, and the availability of data on the HR operational system, we changed to use the number of hours to populate the cube.

The cube thus sums the amount of hours every employee has for each frequency situation of the organization. The number of hours is obtained through the verification of the employee frequency, which is registered in a database that stores the daily amount of hours that the employee works, is absent, or in any other exceptional situation.

Dimensions. Besides some historical dimensions of enrollment data, like workplace, frequency situation and gender, we create a special dimension called Absenteeism.

We considered as Integral Actuation on the company all employees that had worked, or in certain absences in accordance with the Brazilian Consolidation of the Labor Law (CLT, art. 473) [9] during the frequency month.

Absences shorter than thirty days long, and that were not legal ones, were defined as Partial Actuation on the company.

Absences exceeding thirty days long were regarded as Long Term Absence.

4.2 Facts

It's possible to view the amount of hours in a certain frequency situation using only the fact itself, without estimates. The number of hours is split among Integral Actuation (I), Partial Actuation (P) or Long Term Absence (LT).

Besides, we can apply a function to translate the raw number of hours to a generic absenteeism rate, $r(x)$. To get the Integral Actuation rate in the company, we divide the amount of hours of Integral Actuation by the total hours of Actuation, Integral plus Partial, or

$$r(I) = I / (I + P). \quad (1)$$

The absenteeism rate of the Partial Actuation is obtained from the division of the amount of hours of the Partial Actuation by the sum of hours of the Integral Actuation plus the hours of the Partial Actuation, or

$$r(P) = P / (I + P). \quad (2)$$

To get the Long Term Absence rate, just do the division of the amount of hours in Long Term Absence by the sum of hours of the Actuation Integral plus the hours of the Actuation Partial plus the hours in Long Term Absence.

$$r(LT) = LT / (I + P + LT). \quad (3)$$

4.3 Using the Absenteeism Data Mart through a BI software tool

Accessing the Data Mart the managers can create themselves several spreadsheets and charts, or access some pre-built ones. Figure 2 below shows a simple but typical table, where we can perceive the star-schema structure: dimensions on rows and columns, and a fact summarization on the center cells.

Sum(hours)		Month		
Type	Subtype	01/2006	02/2006	03/2006
Actuation	Integral	1540120	1184919	1382770
Actuation	Partial	26176	30929	36842
Long term absent	Long term absent	120522	102414	122611

Fig. 2. Typical BI table, showing dimensions (*gray*) and facts (*white*).

The dimension Time was used for columns on the Month hierarchy level, and the dimension Absenteeism was placed row titles, using both levels Type and Subtype. The fact presented is the number of hours for each situation, summarized by a Sum function. The data were filtered for showing only the months 01/2006, 02/2006 and 03/2006.

A common OLAP tool allows us to apply user-defined functions and calculations to the raw facts. For instance, to the same data used for Figure 2 we can apply the formulas described on section 4.2 for the absenteeism rate on the organization, producing the report shown on Figure 3.

% Abs.		Month		
Type	Subtype	01/2006	02/2006	03/2006
Actuation	Integral	98,3%	97,5%	97,4%
	Partial	1,7%	2,5%	2,6%
Long term absent	Long term absent	7,1%	7,8%	8,0%

Fig. 3. Processing the raw fact using Absenteeism rate formulas.

Since the users' needs for this system concern mainly the Partial Actuation index, we can plot a chart to view its behavior during all the months of 2006, as shows Figure 4, simply changing the filter of the Time dimension and applying the fact data to the Y axis of the chart.

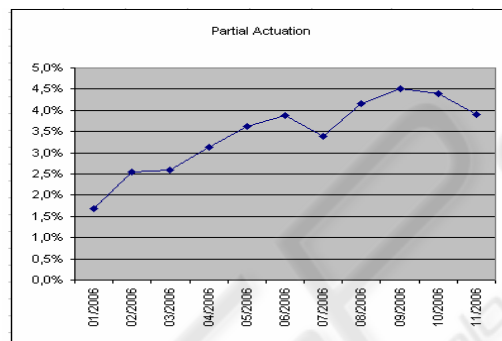


Fig. 4. Chart for the Partial Actuation rate during the year of 2006.

If the user becomes curious about whether this behavior repeats for every year, he or she can **drill-up** the Time dimension, accessing the Year hierarchy level, and producing another report and chart (Figure 5). Managers now can be prevented about workforce needs along the year.

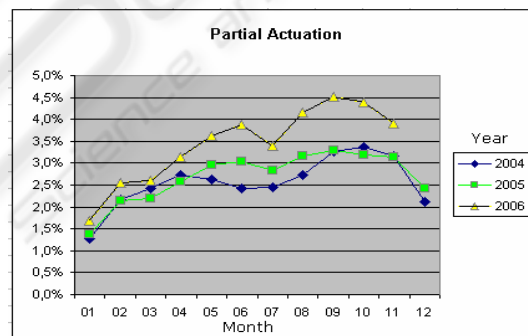


Fig. 5. Drilling-up the Time dimension and showing the last 3 years.

Other dimensions, as Workplace for instance, can be added to the sheet by simply dragging and dropping onto the rows or columns, and then the same procedure of

drilling up and down can be applied for finding departments or offices of the organization where the absenteeism rates require more attention from the Human Resources department.

5 Results Analysis

The first phase of the project resulted on modeling about a dozen reports. The Datmart was loaded with attendance data from the last five years. Some results were presented to managers and HR team. These users provided us a positive feedback and requests for more reports. The generated information was used for real institution's needs, as some strategic decisions related to extra-workforce hiring.

The new reports required by the users were variations of the first ones. To build these reports, few changes on dimensions were enough, for instance, drilling down the workplace dimension or adding/removing function dimension. From all the dimensions provided, the only one that was not requested to be used was the gender dimension.

The choice using the number of hours, on work and absent, as the fact of the cube, shows to be a good approach. The absence of a twelve-hour shift employee and a four-hour shift employee, even in the same working day, are not summarized with the same weight, reflecting the fact that the needs for workforce replacement are not the same for each case.

Regarding the response time needed to obtain a report, a significant improvement was observed, because on the moment of report generation the processing load is light. For instance, one previous report built on the HR operational system that retrieves data on absence situations on the whole organization requires about one hour of processing, due to the amount of data. One report with similar content, that we tried to create using BI tools, runs in a few minutes.

6 Conclusions and Further Work

The described project was started by the need of analysis of the absenteeism for the institution. After a fast survey of the implementation possibilities, was opted to trying a solution using BI. The good feedback received demonstrates that the use of BI for construction of tools for absenteeism management is a good strategy and that the developed model shall evolve.

One of the concerns for the project continuity must be to find ways to explain to all managers, potential users of the system, the very definition of the index of absence to decrease the complexity of the reports interpretation. Nowadays only people who know the concept of the absenteeism index is able to interpret the results correctly. We had experiences with users who don't know this concept and they had made wrong data interpretations.

We also need to deal with the diversity of users' needs regarding the data organization on the dimensions. For instance, once the cube is loaded with absences longer than thirty days as Long Term Absences, users are not able to change

themselves the duration threshold to fifteen days. This need requires the creation of a new cube and dimensions, a technical task.

Next phases of the project must address Occupational Health needs. We are planning to add a dimension that allows the user to view the diseases related to the International Classification of Diseases (ICD) [10]. This solution is easy to implement due to the hierarchical nature of ICD codes, which fits well to a DM dimension.

The Star Schema storage provides a good compromise between data availability and confidentiality, as no identification of individuals can be accessed from the OLAP, only summarized data, being impossible to relate diseases to persons, but otherwise providing a good landscape of professional sickness over the organization geography, time and occupations.

BI tools empower the user's view of the facts. The dynamic character of report generation is helpful for the absenteeism study, since the causes of absence are not always predictable, and neither is the search for the solutions.

Acknowledgements

The authors thank Unicamp/DGRH, Unicamp/AFPU and Unicamp/PRDU, especially Prof. Dr. Paulo Eduardo Moreira Rodrigues da Silva, for financial support, and Sueli Bonilha Esteves, Gláucia Beatriz de Freitas Lorenzetti and Regina Bernardo da Luz as project sponsors.

We also thank Telma Freitas and the anonymous reviewers, who provided important feedback about this work.

References

1. Chiavenato, I. Recursos Humanos. Ed. Compacta, 7. Ed. - Sao Paulo (2002)
2. Quick, T. C. and Lapertosa, J. B. Analise do absentismo em usina siderurgica. Revista Brasileira de Saude Ocupacional. Sao Paulo, 10(40): 62-67, (1982).
3. Couto, H Como evitar o absentismo e elaborar um indice adequado. Revista tendencia do trabalho, Administracao de Pessoal (1998).
4. Sipageh project. Available online on http://www.economicas.unisinos.br/sipageh/_indic/_absenteismo.php. Accessed on 02/2007
5. Bradford Factor. <http://www.incomesdata.co.uk/studies/bradford.htm>. Accessed on 02/2007
6. Power, Daniel J.: Decision Support Systems: Concepts and Resources for Managers. Quorum Books (2002).
7. Inmon, William, H.: Building the Data Warehouse. Wiley Publishers; 4th edition (2005).
8. Wikipedia, the free encyclopedia, http://en.wikipedia.org/wiki/Data_warehouse. Accessed on 02/2007.
9. CLT, the Brazilian Consolidation of the Labor Law. Available online (in Portuguese) on https://www.planalto.gov.br/ccivil_03/decreto-lei/Del5452.htm Accessed on 02/2007.
10. ICD, International Statistical Classification of Diseases and Related Health Problems. Available online on <http://en.wikipedia.org/wiki/ICD>. Accessed on 02/2007.