PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

UNIVERSITY IBN KHALDOUN - TIARET

MEMORY

Introduced to:

FACULTY MATHEMATICS AND COMPUTING
COMPUTER DEPARTMENT

For graduation from:

MASTER

Specialty: *Software Engineering*

By:

BOUABDELLI Abdelaziz and ABDI Abdallah

On the subject

A Decision-making System for The Management of Supermarkets

Publicly supported on .. / 07/2019 in Tiaret before the jury composed of:

Mr. CHIKHAOUI Ahmed	Dr.	University of Tiaret	President
Mr. BAKKAR Khaled	MAA	University of Tiaret	Framer
Mr. OUARED Abdelkader	Dr.	University of Tiaret	Co-Framer
Mr. HATAB Nouredine	MAA	University of Tiaret	Examiner

Dedications

I dedicate this modest work

To my parents

That I will not stop loving, Who fed me with their love, their joy of living, and who sacrificed themselves to make me what I am today.

To my brothers and my sister

Who have been unequivocal support.

To my wife

Who has always encouraged me and to whom I owe a lot to my success.

To my son Ayoub

The light of my life

To all my family

Who have been present for me.

To my partner

Abdelaziz.

To my friends

Youcef, Saad and Kaddar with whom I spent years unforgettable full of humor, joy and madness.

To all my colleagues at work (Department of the Environment)

Abdallah.

Dedications

I dedicate this modest work

To my dearest parents

They who offered me everything they never had, who sacrificed themselves constantly for my happiness. I thank those who have always listened to me and who have enlightened me with their valuable advice.

My sisters

Who have always been by my side.

To my wife

With whom I wish to spend a life full of joy and happiness.

To all my family

Who have been present for me.

To my partner

Abdallah.

To my friends

Youcef, Saad and Kaddar with whom I spent years unforgettable full of humor, joy and madness.

Abdelaziz.

thanks

It is with great pleasure that I reserve this page, as a sign of gratitude and gratitude to all those who helped me to carry out this work.

I thank, first of all, **Mr Bakkar Khaled** for his rigor and the relevance of his judgments which were very constructive and allowed me to do this work.

I am also grateful for his contribution to the judicious improvement of the quality of this document.

I thank **Dr Ouared AEK** for his meticulous co-supervision and his good pedagogic and scientific qualities. His wise advice was invaluable and his kindness unequaled.

I thank **Dr Chikhaoui Ahmed** for having done as the honor of being president of the jury,

I thank cordially **Mr Hatab Nouredine**, for having accepted to be examiner of this Memory.

To all members of Tiaret University and in particular the computer science department.

I would like to express my gratitude to my relatives : my dear parents, my brothers and my sisters, my wife .

Without their support, trust and encouragement, I would never have done it.

To my friends and colleagues without exception, for their presence, their respect and their kindness.

And finally, thank you to everyone who participated in this work.

BOUABDELLI Abdelaziz and ABDI Abdallah.

Contents

List of figures xv

List of tables xvii

General Introduction 1

0.1 General Introduction . . . 1
0.2 Problematic . . . 1
0.3 Objectives . . . 2

Partie I Bibliographic synthesis 3

Chapitre 1 Decision systems 5

1.1 Introduction . . . 5
1.2 Definition of decision-making systems . . . 6
1.3 Use of the decision-making system . . . 6
1.3.1 Top Management (need for strategic management) . . . 7
1.3.2 Middle management (need for operational management) . . . 7
1.3.3 Business actors (standard needs) . . . 7
1.3.4 Experts (special needs) . . . 7
1.4 History of the decision-making system . . . 7
1.5 The features of the decision-making system . . . 8
1.6 Decision system architecture : . . . 8
1.6.1 The sources of data : . . . 9
1.6.2 Data warehouses and Data Mart's : . . . 9
1.6.3 Analysis and reparation tools : . . . 10
1.7 Conclusion : . . . 10

Chapitre 2 Data warehouses **11**

2.1 Introduction : . . . 11
2.2 Definition of a Business Intelligence : . . . 12
2.3 Definition of a Data warehouse : . . . 14
2.4 History of the Data warehouse : . . . 15
2.5 Objectives of the Data Warehouse : . . . 16
2.6 Data Structures of a Data Warehouse : . . . 16
2.7 The elements of a Data Warehouse : . . . 17
2.7.1 Source system : . . . 17
2.7.2 The data preparation area : . . . 18
2.7.3 The data presentation area : . . . 18
2.7.4 Data access tools : . . . 18
2.8 Data Warehouse Architecture : . . . 19
2.8.1 The "back room" : it allows : . . . 19
2.8.2 The front room : it allows : . . . 19
2.9 Data Warehouse vs. Databases : . . . 20
2.10 Data Modeling of the Data Warehouse : . . . 20
2.10.1 Dimensional modeling : . . . 20
2.10.1.1 Tables of facts and tables of dimensions : . . . 20
2.10.2 Different models of dimensional modeling : . . . 21
2.10.2.1 Star model (Star Diagram) : . . . 22
2.10.2.2 Flake Model (Snow Flake Pattern) : . . . 22
2.10.2.3 Constellation model : . . . 22
2.11 The data cube : . . . 23
2.12 The OLAP concept : . . . 23
2.12.1 OLAP storage architectures : . . . 23
2.12.1.1 M-OLAP (Multidimensional OLAP) : . . . 24
2.12.1.2 R-OLAP (OLAP Relational) : . . . 24
2.12.1.3 H-OLAP (OLAP Hybrid) : . . . 25
2.12.2 Comparison between MOLAP, ROLAP and HOLAP signed in Table 2.2 : . . . 25
2.13 Navigation in the data : . . . 25
2.13.1 Set-up operations "Slice Dice" : . . . 26
2.13.1.1 Slicing . . . 26
2.13.1.2 Dicing . . . 26
2.13.2 Roll-up Drill-down granularity operations : . . . 27

2.13.2.1 The Drill-down : . . . 27
2.13.2.2 The Roll-up : . . . 27
2.14 Conclusion : . . . 28

Chapitre 3 Conduct and construction of the Data Warehouse **29**
3.1 Data Warehouse Building Methodologies . . . 29
3.1.1 Top-Down . . . 29
3.1.2 Bottom-Up . . . 29
3.1.3 Top-Down vs Bottom-up . . . 30
3.1.4 Middle-Out . . . 30
3.2 Data Warehouse Power Supply . . . 31
3.2.1 Extract-Transform-Load ($\mathcal{ETL}$) process . . . 31
3.2.1.1 Data extraction . . . 31
3.2.1.2 Data transformation . . . 32
3.2.1.3 Loading data . . . 32
3.2.1.4 Data Profiling and Data Quality Control . . . 32
3.3 The deployment and implementation of a Data warehouse . . . 32
3.3.1 Reporting . . . 32
3.3.2 Dashboard . . . 33
3.3.2.1 The objectives of the dashboards . . . 34
3.3.2.2 Design of a dashboard . . . 34
3.4 Data Mining . . . 35
3.5 Data mining techniques . . . 35
3.5.1 Statistical techniques . . . 35
3.5.2 Learning techniques . . . 35
3.6 Definition of Multiple linear regression . . . 36
3.7 Conclusion . . . 36

Partie II Preliminary study **37**

Chapitre 4 Presentation of the host company , Study of the existing system and Study of needs **39**
4.1 Definition Supermarket . . . 39
4.2 Introduction . . . 40
4.3 Operating systems . . . 40
4.4 Decision Support Systems . . . 40

4.5 Computer applications and licenses . . . 42
4.6 Conclusion . . . 43
4.7 Introduction . . . 43
4.8 Presentation of information gathering approaches . . . 43
4.9 Presentation of the approaches used . . . 45
4.10 Obstacles encountered . . . 45
4.11 Needs analysis . . . 45
4.12 Identification of users . . . 47
4.13 Conclusion . . . 47

Partie III Design and Implementation of the project 49

Chapitre 5 Design of the zoned storage 51
5.1 Introduction . . . 51
5.2 Participating dimensions : . . . 52
5.3 Process modeling . . . 52
5.3.1 Modelization of process " Sale " . . . 52
5.3.2 Modeling of the "Purchasing" process . . . 57
5.3.3 Modeling of process " Stock " . . . 59
5.4 Conclusion . . . 61

Chapitre 6 Design of the zoned Power 63
6.1 Introduction . . . 63
6.2 Studies of the sources of data . . . 63
6.3 Data Warehouse power supply . . . 64
6.4 Conclusion . . . 65

Chapitre 7 Design of the cubes dimensional 67
7.1 Introduction . . . 67
7.2 levels and hierarchies of the dimensions . . . 67
7.3 The listing of the cubes . . . 68
7.4 Presentation of the cubes dimensional . . . 69
7.5 Conclusion . . . 71

Chapitre 8 implementation 73
8.1 The architecture of the solution . . . 73

8.2 Resources used . . . 75
8.2.1 Material resources . . . 75
8.2.2 Software resources . . . 76
8.2.3 Comparative study of Open Source BI tools . . . 76
8.3 Implementation of the solution . . . 77
8.3.1 Implementation of the ETL . . . 77
8.3.1.1 Definition of connections to data sources . . . 77
8.3.1.2 Loading the "analysis" database allows you to . . . 77
8.3.1.3 Cleaning of the "analysis" data Area . . . 78
8.3.1.4 Loading the DW . . . 78
8.3.2 Implementation of the data analysis and reporting component . . . 79
8.3.2.1 Creating the OLAP schema . . . 79
8.3.2.2 Data analysis and restitution component . . . 79
8.3.2.3 Dashboard . . . 81
8.3.2.4 Regression equation . . . 81
8.3.2.5 The modelling processes . . . 81
8.3.3 Choice of method . . . 82
8.4 The multiple linear model . . . 82
8.4.1 Writing the model . . . 82
8.4.2 Principle of the calculation . . . 82
8.5 Conclusion . . . 85

Chapitre 9 Overall conclusion **87**
9.1 Overall conclusion . . . 87

Partie IV Abriviation and Bebliography **89**

List of figures

1.1 Users of the decision-making system 6
1.2 Architecture of decision-making systems 9

2.1 demonstration of Business intelligence 12
2.2 how BI it works? 13
2.3 causes of using BI 14
2.4 Subject orientation of the Data Warehouse 14
2.5 Integration in the Data Warehouse 15
2.6 The Data Warehouse Historization 15
2.7 Data structure of a Data Warehouse 17
2.8 Data mart in the Data Warehouse 18
2.9 Designing a Data Warehouse 19
2.10 A fact table 21
2.11 A dimension table 21
2.12 A star schema 22
2.13 A flake diagram 22
2.14 A data cube 23
2.15 Principle of the MOLAP architecture. 24
2.16 Principle of the ROLAP architecture. 25
2.17 Example of slicing. 26
2.18 Example of Dicing. 26
2.19 Example of Drill-Down. 27
2.20 Example of Roll-Up. 28

3.1 DW architecture according to Inmon. 29
3.2 Kimball's DW architecture. 30

3.3 The ETL process. 31
3.4 Example of an output status in reporting. 33
3.5 Example of an output status in reporting. 34

4.1 Use of a SAGE report to obtain turnover. 41
4.2 Choice of valuation and period. 42

5.1 Star model of the "Sale" process. 57
5.2 Star model of the "Buy" process. 59
5.3 Star model of the "Stock" process. 60

6.1 Architecture of the power solution. 64
6.2 Descriptive diagram of the DW feeding process. 65

7.1 Dimensional Cube "Sale". 69
7.2 Dimensional Cube "Purchase". 70
7.3 Dimensional Cube "Stock". 70

8.1 Overall solution architecture. 74
8.2 The architecture of the restitution part. 75
8.3 Comparison table of the best open source BI tools. 76
8.4 Connection to data sources. 77
8.5 Loading analysis Area. 78
8.6 ETL. 79
8.7 Example of purchase analyses (total amount to be paid to suppliers by product). 80
8.8 Example of a graph (the amount of the remaining merchandise to be paid to suppliers). 81
8.9 Dashboard of the physical achievements of the forecast 84
8.10 Dashboard of the physical achievements of the promotion(tendence). 84
8.11 Dashboard of the physical achievements of the promotion. 85

List of tables

2.1 Comparison between Data Warehouse and BDD 20
2.2 Comparative table between R-OLAP, M- OLAP and H-OLAP. 25

3.1 Comparative table between the approaches of setting up a Data warehouse according to Inmon and Kimball. 30

4.1 Summary table of needs detected. 47

5.1 The dimensions participating in the different components. 52
5.2 Descriptive table of the "Date" dimension. 54
5.3 Descriptive table of the "Product" dimension. 55
5.4 Descriptive table of the "Deposit" dimension. 55
5.5 Descriptive table of the "Customer" dimension. 56
5.6 Descriptive table of the "Point of sale" dimension. 56
5.7 Descriptive table of the "Invoice" dimension. 56
5.8 Listing of the aggregates of"Sale". 57
5.9 Descriptive table of the "Provider" dimension. 58
5.10 Descriptive table of the "Order" dimension. 58
5.11 List of "Buy" aggregates. 59
5.12 Descriptive table of the "type of stock" dimension. 60
5.13 List of aggregates of the "Stock". 61

7.1 List of levels and hierarchies of each dimension. 68
7.2 List of multidimensional cubes. 69

8.1 Tools used during the development of the solution. 76

General Introduction

0.1 General Introduction

At all levels of the company, employees make decisions to improve business performance and to cope with a highly competitive market that evolves with extreme speed. For this the company needs to better understand and to satisfy its customers. To do this, policymakers need to make informed decisions and based on the clear, reliable and relevant information. The business problem is to analyze a large volume of data stored in operational systems. These systems operational management of the day-to-day business (accounting, production, etc.) more over the data are stored imperfectly because of conventional means, which makes their exploitation and exploration difficult or impossible. Therefore, companies need a tool that can meet these requirements, and that allows to explore these astronomical amounts of data, and to analyze and collect the information needed to make the decisions. This is the case of Business Tools Intelligence that allows the company to see a perfect image of its business and facilitate informed decision-making and identify shortcomings in its functions and services.

In recent years, after the wave of deployment of transactional systems (the s ERP), will have been those of decision-making tools. Analysts are more over just about all unanimous to foresee a strong progression of their use in the years to come. Yesterday technical tools, complex and expensive to set up, their deployment has become simpler and their use has become democratized. Some ERP editors have even integrated their software a decision offer sold as "immediately" operational. This democratization has gone hand in hand with an explosion of features. [1]

Today, a decision-making system is at the top of the list of technology tools of the supermarkets, and that as decision support. In order to position itself on the market Algerian. To this end, the company has entrusted us with the responsibility of designing and producing such a system as part of our end-of-term project entitled "Design and Implementation of a decision-making system for the stock, supply and sale activity within the Supermarkets ".

0.2 Problematic

Supermarkets is a company specializing in marketing and assist all needs of human as (food and cosmetics and pharmaceuticals, and clothing and all DIY tools).

1. An ERP (Enterprise Resource Planning) or also called ERP (Integrated Management Software Package) is an information system that allows you to manage and monitor daily all the information and operational services of a company.

Supermarkets managers know that good management is essential for the company can follow its annual planning and thus guarantee the achievement of objectives according to the strategy it has set itself.

And that by effectively using the data stored in the database to provide you with relevant information and to decision makers the information needed to facilitate decision-making.

For this a powerful tool is necessary to meet the strategic expectations, tactics and operations of the company. And face the daily difficulties concerning the analysis and dissemination of data.

The current system is encountering some difficulties that we have seen at different levels in our study, and that we present them as follows :

1. Difficulty and slowness in decision-making due to lack of means of analysis of commercial activity.
2. Information provided to managers is not information relevant because of the misuse of the data.
3. The absence of performance indicators for commercial activity.

   "" So,with all this problem have we the solution ?

0.3 Objectives

In order to solve the problems encountered at the level of Supermarkets, we were entrusted design and implement a decision support system for commercial activity (supply, stock and sale).

And this, by implementing a Data Warehouse that will provide the necessary information that provides reliable support to decision-makers for better decision-making.

To do this, we have set the following objectives :

— Facilitate the exploitation of stored information (data organized according to different dimensions).
— Combine and cross the previously unrelated information so to obtain a complete view and an immediate analysis of the information.
— Propose and implement performance indicators for sales services, stock and supply.
— Gather the necessary elements for decision-making through tables of on board and reduce decision time cause time it's money.
— Propose alerts with a view to anticipating stock outs, and going beyond delivery delay.

Part I

Bibliographic synthesis

Chapitre

1 Decision systems

1.1 Introduction

In an environment that has become increasingly uncertain and complex, it is difficult for companies to apprehend it because of the diversity of sources of information. And market globalization, the company has to adapt, if possible, to anticipate, sometimes influence. The problem is no longer so much to acquire a mass of data, but of exploit it.

However, when it comes to organizational decision marking ,perhaps you would n't want to take risk, the reason are obvious one wrong decision can make many things go awry , including you brand image , product life cycle, financial standing and employer band .

In many situations , you're unable to apply fundamentals if economics ,Statistics and operations research to make lucid choice.

So , you need some knowledge based System that support business decision making activities ,this is when a decision Support system Comes into picture. "it is a computer Based System that helps you make planning, manufacturing ,operations and management decision , based on information available.

Say, for **example** , you need to devise a supply chain movement Strategy . **how will you do this ?** A ($\mathcal{DSS}$) analyzies the stock of inventory and production movement. [2].

Basis the data available, it compare the out comes of different decision, helping you figure out what may work best currently , it helps you establish a supply chain movement that works.

Let's take another **example**, suppose you want to optimize your sales, **how will you do it ?** Obviously, you'll make some assumptions, collect data, facts and figures, generate reports, examine patterns and finally make a decision . when you do it manually, you never know what mistakes you're committing, a DSS in such a situation , can gather and analysis data and make predication by monitoring existing pattern , it speeds up the whole process, giving you in sights into how you can optimize your sales process.

Faced with this need, publishers have developed solutions to help managers. Their offers, grouped under the term decision-making systems, are today very rich but also very different in terms of functionality and approach.

2. https://fr.wikipedia.org/wiki/Informatique_d

1.2 Definition of decision-making systems

What does "decision making" mean ? The dictator "Oxford Advanced Learner" gives the following definition : "the process decision about something important, especially in a group of people or in an organization " [**Oxford, 2005**].

Where does the term " decision-making system " come from ?

($\mathcal{SID}$) is a system that allows business decision-makers to have information relevant and powerful analytical tools to help make the right decisions for the right moment " [**Devisy 02**].

Decision systems are a set of technologies that allow to gain access and understanding of pilotage data more quickly, so that they make better and faster decisions to reach the objectives of their organization. It involves collecting internal and external data, transform into information that will then be analyzed by users for the purpose of make better decisions and improve business performance.

1.3 Use of the decision-making system

The people who benefit from and use the business decision-making system are the decision-makers, who are usually "marketers" or analysts who make plans marketing that allows them to better target their customers and retain them then they have synthesized information and summary data of their activities for to follow well. Depending on the needs, there are four classes of workers who use the decision-making system as we represent in the **figure1.1**

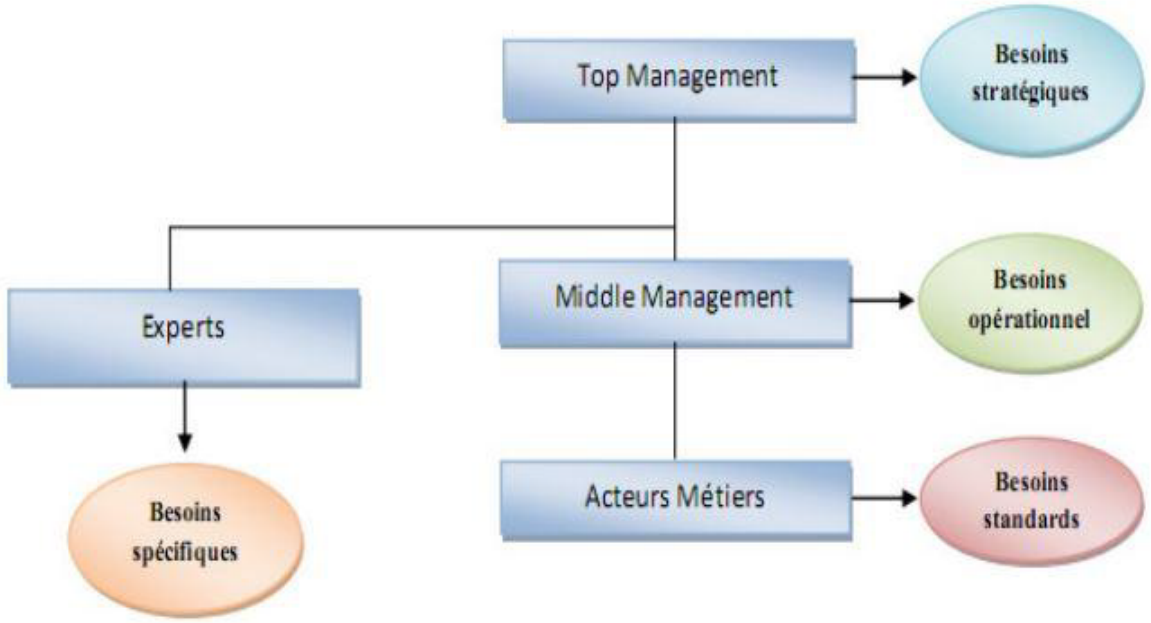


Figure 1.1 – Users of the decision-making system

1.3.1 Top Management (need for strategic management)

Their first concern is the management of the performance by the follow-up of the objectives of the organization and consolidation of operational indicators. The decision-making system offers them information synthesized in the form of pilot indicators presented via dashboards.

1.3.2 Middle management (need for operational management)

Operational management concerns the daily monitoring of operational processes, and allows in-depth analysis of deviations. The user has the opportunity to "navigate" and to "zoom in" on the data in order to analyze in depth the situation present.

1.3.3 Business actors (standard needs)

They have operational activities and have standard needs known in advance ; Those are consumers of information only, who are generally content to predefined static reports.

1.3.4 Experts (special needs)

These are end users with skilled trades requiring skill advanced (finance, marketing ...) ; they have specific needs in their trades (budget consolidation, customer segmentation ...), also, the decision-making system must offer them powerful tools to effectively meet their expectations.

1.4 History of the decision-making system

Here is a brief summary of the steps that have marked the march of the concept of decision-making system : **[Burquier 09]**. [3]

In 1962 Ken Iverson published the first multidimensional language (APL A Programming language), which has been translated by a first multidimensional tool aimed at marketing type applications. Ten years later, the OLAP engine designed for financial applications (Banks, insurance ...).

In 1993 E. Codd dictated the universal rules that describe the OLAP engines (model relational). Until the two great theoreticians **Ralph Kimball** and **Bill Inmon** have completely upset the world of information by turning data into a Reliable and consistent information in business . What pushed the big firms IT like MICROSOFT, IBM create powerful analytics tools for setting in application of business-oriented data. In addition, the competition forced them to produce more tools and software to implement the strategic indicators of the company and how to interpret them to set forecasts and detect targets or trends.

In 2000, Microsoft did a real facelift of business intelligence and wanted gaining a foothold in the DBMS market with Oracle and IBM, the world leader in business Intelligence, by creating the SQL Server

3. Bertrand Burquier, "Business Intelligence with SQL Server 2008- Implementation of a decision-making project", Editions DUNOD 2009.

2000 version. This led to competition between firms and especially between MICROSOFT and IBM has advanced in a way spectacular BI and its applications.

1.5 The features of the decision-making system

Users' needs can be grouped into 4 main categories : Simulate, Analyze Data, Produce management reports, follow and control. In response to these needs, decision-making system offers the following features :

Simulate : Management of calculation models (Automatic computation of complex data set depending on : parameters entered by users and management rules).

Example of use : Business plan development Collaborative development (Adding work flow features to templates simulation described previously).

Example of use : budgeting.

Analyze the data :

OLAP features. Establishment of dynamic analyzes multi-dimensional with the ability to sort, filter zoom within data.

Example of use :determination of causes explaining the evolution of a summary data (turnover).

Produce management reports : Reporting functionality. Snowshoes to produce in a simple way and fast tables of data incorporating more or less sophisticated.

Example of use :production of management status on demand.

Follow and control :

Development of dashboards : Production and automatic distribution to regular frequency of board charts gathering heterogeneous data.

Example of use : Production of graphical dashboards destination of operational managers (sales managers).

Alerting :

Example of use : alert on finding a payment incident. It is rare that all these features are implemented in a business. The implementation s implemented are also often performed by functional area (sales, purchases ...).

1.6 Decision system architecture :

On a practical and technical level, the decision-making system consists of a family of tools information systems and software packages ensuring the functioning of the information. It proposes to use the data

transiting through the Information System, production data more often than not, in information that could be exploited for decisional purposes.

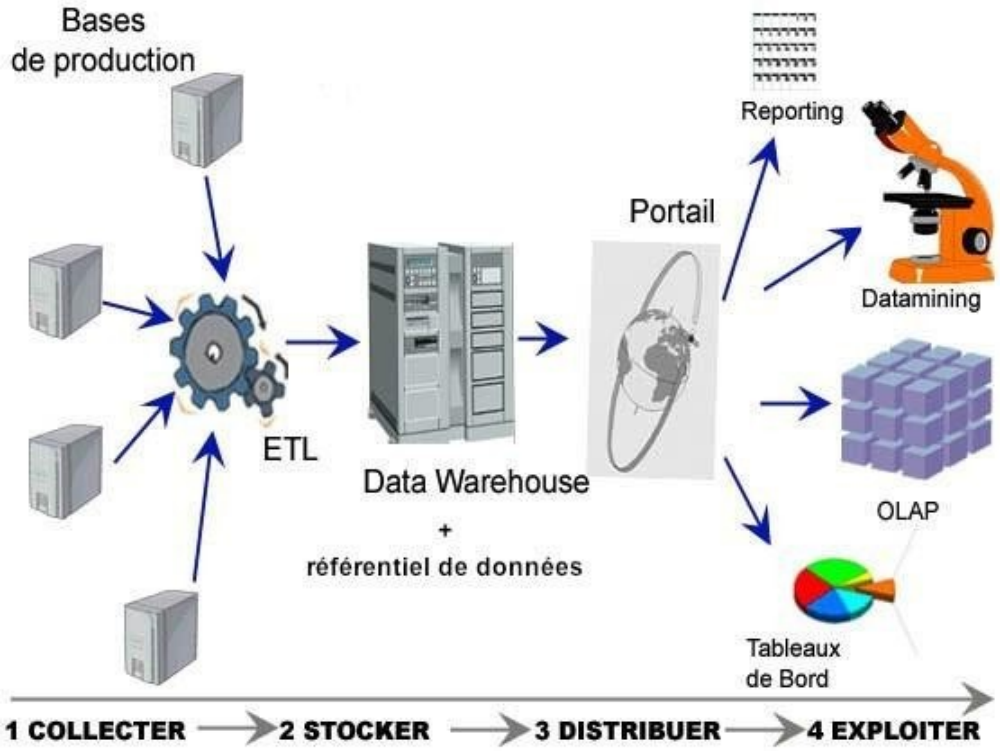


FIGURE 1.2 – Architecture of decision-making systems

The architecture of a decision-making system, shown in **Figure 1.2**, includes three main components :

1.6.1 The sources of data :

Collecting and integrating data stored in various primary sources and secondary, which are heterogeneous, is the first step. Most of data belonging to operational systems, but may also include unstructured documents. In general, a major effort is needed to unify and integrate the different sources of data.

1.6.2 Data warehouses and Data Mart's :

Production databases can not be used for a farm decision. The raw data are not ready for this purpose and the requests decision makers are particularly greedy in machine resources. Data, previously cleaned and consolidated, will be stored in a database specialized : The data warehouse or Data Mart.

The data warehouse is the heart of the system it contains data from different sources. The use of tools extractions, transformations and shipments, known as **ETLs**, will be necessary to load the data into the data warehouse.

1.6.3 Analysis and reparation tools :

Once the data is stored, cleaned, consolidated and accessible, it can be used. Depending on the needs, different types of mining and mining tools will be considered.

1.7 Conclusion :

In recent years, decision makers have greatly improved. They have become easier to implement, use and exploit, their integration with transactional applications.

But the risks of not taking full advantage of its decision-making solution and limiting itself to administrative productivity gains related to improving the dissemination of information piloting are real.

In the next chapter , we will treat **the Data Warehouse** corresponding to the exploitation company data, to facilitate decision-making.

Chapitre

2

Data warehouses

Sommaire

2.1 Introduction : . . . 11
2.2 Definition of a Business Intelligence : . . . 12
2.3 Definition of a Data warehouse : . . . 14
2.4 History of the Data warehouse : . . . 15
2.5 Objectives of the Data Warehouse : . . . 16
2.6 Data Structures of a Data Warehouse : . . . 16
2.7 The elements of a Data Warehouse : . . . 17
2.7.1 Source system : . . . 17
2.7.2 The data preparation area : . . . 18
2.7.3 The data presentation area : . . . 18
2.7.4 Data access tools : . . . 18
2.8 Data Warehouse Architecture : . . . 19
2.8.1 The "back room" : it allows : . . . 19
2.8.2 The front room : it allows : . . . 19
2.9 Data Warehouse vs. Databases : . . . 20
2.10 Data Modeling of the Data Warehouse : . . . 20
2.10.1 Dimensional modeling : . . . 20
2.10.2 Different models of dimensional modeling : . . . 21
2.11 The data cube : . . . 23
2.12 The OLAP concept : . . . 23
2.12.1 OLAP storage architectures : . . . 23
2.12.2 Comparison between MOLAP, ROLAP and HOLAP signed in Table 2.2 : 25
2.13 Navigation in the data : . . . 25
2.13.1 Set-up operations "Slice Dice" : . . . 26
2.13.2 Roll-up Drill-down granularity operations : . . . 27
2.14 Conclusion : . . . 28

2.1 Introduction :

A data warehouse ($\mathcal{DW}$) is a relational or multidimensional database that is designed for query and analysis. They are not optimized for transaction processing, which is the domain of OLTP systems.

Data warehouses usually consolidate historical and transactional data derived from multiple sources. Data warehouses separate analysis workload from transaction workload and enable an organization to consolidate data from several sources.

Data in a data warehouse usually stores many months or years of data to support historical analysis. The data in a data warehouse is typically loaded through an extraction, transformation, and loading (**ETL**) process from one or more data sources such as OLTP applications, mainframe applications, or external data providers.

End users of the data warehouse perform data analyses that are often time-related. Examples include consolidation of last year's sales figures, inventory analysis, and profit by product and by customer. More sophisticated analyses include trend analyses and data mining, which use existing data to forecast trends or predict futures. The data warehouse typically provides the foundation for a business intelligence environment.

This guide covers relational implementations, including star schemas. See **Oracle Database Data Warehousing Guide** for more details regarding multidimensional data warehouses. The DW is a database of all the functional data of a business. It falls within the framework of business intelligence.

the concept of DW appeared in the early 90s to transform and integrate data from different DBs to increase the productivity of the company. In this chapter, we will present the need and architecture of Business intelligence and the definition and the construction as well as the fundamental concepts of a data Warehouse.

2.2 Definition of a Business Intelligence :

($\mathcal{BI}$) is the activity which contributes to the growth of any company or in other words is the act of transforming row / operational data into useful information for business analysis as we see it in the **Figures2.1**.

FIGURE 2.1 – demonstration of Business intelligence

How does it Work ?

1. BI based on Data Warehouse technology **Extracts** information from a company's operational systems.

2. The data is **transformed** (cleaned and integrated), and **loaded** into Data Warehouses.

3. Since this data is credible, it is used for business insights.

like we mentioned in the **Figures 2.2** :

FIGURE 2.2 – how BI it works ?

But Why data Warehousing ?

Why Data warehouse ?

— Data collected from various sources stored in various database cannot be directly visualized.
— The data first needs to be **integrated** and then **processed** before visualization takes place.

All the reason to use data warehouse is in the **Figures2.3**.

FIGURE 2.3 – causes of using BI

2.3 Definition of a Data warehouse :

"The Data Warehouse is a collection of topic-oriented, integrated, non-proprietary data volatile and historized, organized to support a decision-making process. " [**Bill Inmon (1996)**].

There are five key features for the Data Warehouse :

Subject-oriented : Data is classified and stored by subject and not by application job as representation in **Figures2.4**.

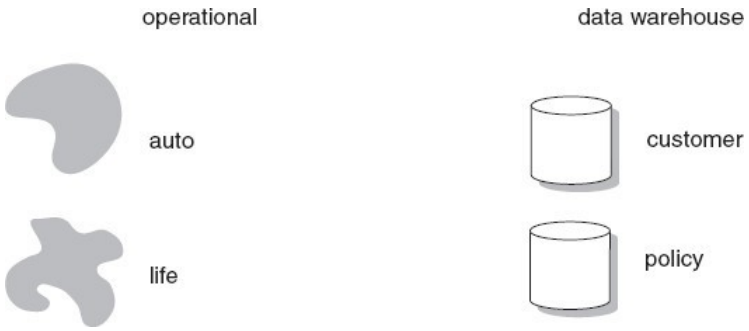


FIGURE 2.4 – Subject orientation of the Data Warehouse

Integrated : Data on a particular subject resident in more than one database will be embedded in a Data Warehouse and any inconsistencies and anomalies in the data must be maintained before loading this data into the Data Warehouse.

We collect the information to resume it int the **Figure 2.5**.

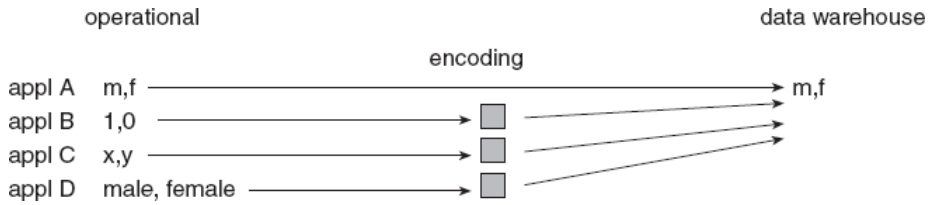


FIGURE 2.5 – Integration in the Data Warehouse

Non-volatile : DW data can not be deleted or updated. This is different from logic Transactional databases.

Historized : Data from a Data warehouse is scalable over time and does not contain typically, no transaction data in real time and we mentioned it in the **Figures 2.6**.

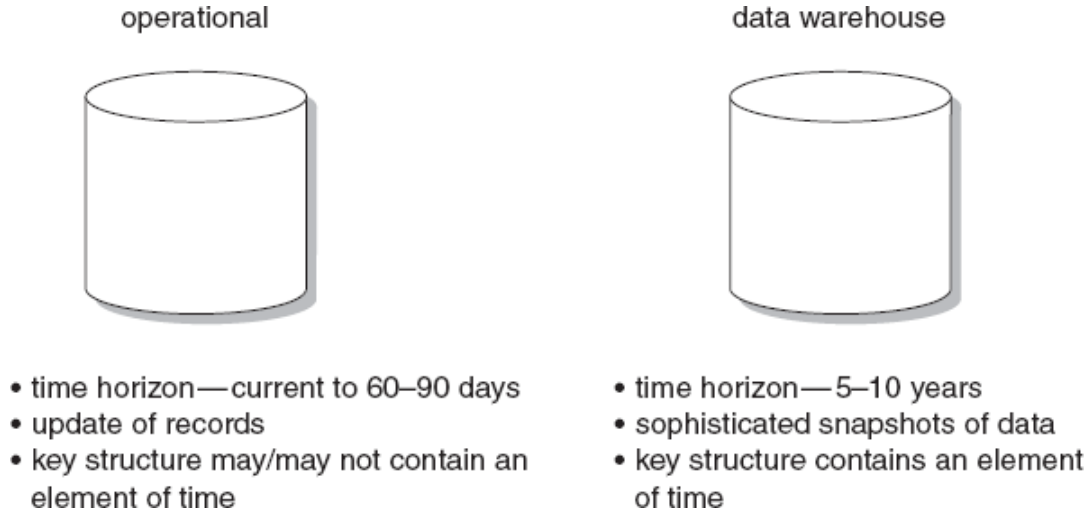


FIGURE 2.6 – The Data Warehouse Historization

Organized to support a decision-making process : Data Warehouse data is sorted and stored to enable the execution of decision support processes (Reporting, Data Mining ....).

2.4 History of the Data warehouse :

There are several concepts before the current concept of data warehouse. They had all purpose of allowing the company to have a source of information relevant decisions. We will quote :

The information center : The first solution was to share access and use of the same resources used by the operand system (BDD for management and treatment).The idea of creating a production database copy and assigning one to each, which allows the isolation of the production process from that of the analysis. Duplication data was being made every day, every week or even every month. This solution could not last long because the analysts with their requests large volumes have blocked the operation of the operational systems, which causes major troubles in the company.

Executive Information System : This solution gave executives the impression of controlling the activity of their companies. This concept proposes a permanent dashboard of key indicators of the company. But the display of a graph on his post seemed simple to him, he supposed a complex mechanics in am have feedback and aggregation of information.

Data warehouses : After several years of use of the infocentre. The Data Warehouse then became a new source of information, provided with collected and consolidated data different internal and external sources.

2.5 Objectives of the Data Warehouse :

"The needs, the customers, the structures and the rhythms of the Data Warehouse are profoundly different from those of operational systems ". **[Kimball, 2002]**.

The objectives of Data Warehouse are :

Make information easily accessible : The contents of the data warehouse must be easily understood. The data must be meaningful and their significance obvious to decision-makers and to consultors.

Make information consistent : Warehouse data must be credible and assembled from different sources of the organization and cleaned.

Constitute an adaptable and resilient source of change : The **DW** is conceived in a perspective of perpetual modifications of its environment. The arrival of new questions should not upset either the data and the existing technologies.

Presents a safe stronghold protecting the organization's data : The most valuable information from an organization is kept in the warehouse of data, it must therefore effectively control access to this confidential data.

Constitute the decision-making base of the company : The DW produces only one output "the decisions made by the reality base that it reveals. These decisions are the added value of the DW.

The acceptance of the DW by analysts and decision-makers : There is no need to build an elegant solution using the best products and platforms, if the business community has not adopted and continues use it.

2.6 Data Structures of a Data Warehouse :

A DW is structured in two axes, each axis is structured to two classes of data. This structure is defined by Inmon **[Inmon, 2005]** as follows in **Figure 2.7**

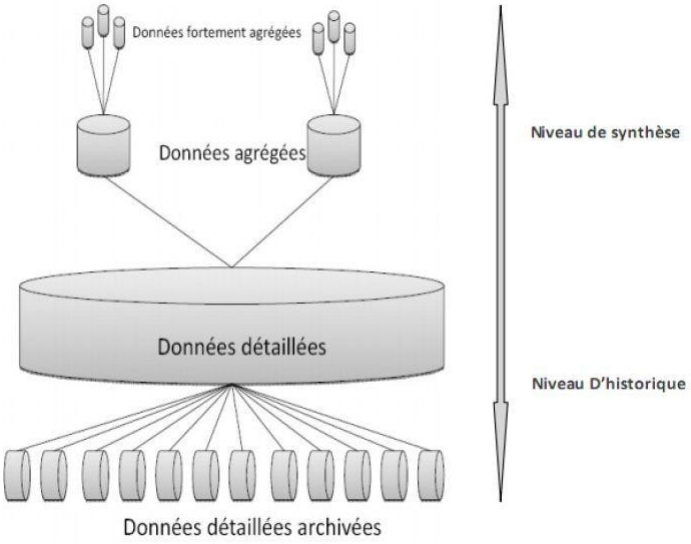


FIGURE 2.7 – Data structure of a Data Warehouse

Historical axis : The volumes to be processed and handled in this area are larger than those managed in transactional.

- Detailed data : they reflect the most recent events, frequently accessed, the volumes to be processed are more important than those managed in transactional by which they are detailed.
- Detailed archived data : old data rarely requested, stored in a mass storage disk, inexpensive, to the same level of detail than the detailed data.

Synthetic axis : The data correspond to already calculated analysis elements representing the user needs.

- Aggregate data : in this case the information is composed for example of the sum of purchases and unit (per month, per product, ...).
- Highly aggregated data : they are almost like data aggregated, they have a huge impact on performance, in some cases they can execute queries 100 times, or 1000 times faster.

2.7 The elements of a Data Warehouse :

A Data warehouse consists essentially of four components : source system, data preparation area, data presentation area and data access tools.

2.7.1 Source system :

A Data Warehouse typically groups data from multiple sources. The sources are external to the Data Warehouse. The goal is to gather the necessary data without affecting the operation of the source systems. These sources can be :

- **Heterogeneous sources :**unstructured document like Word documents, flat files and spread-

sheets.

— **Production databases :** which can be many, different and relocated geographically.
— **external files :** files can be acquired from companies specialized in building and reselling data.

2.7.2 The data preparation area :

In a way is the DW workshop. It consists of a set of processes called ETL, "Extract, Transform and Load". This is where the raw data is loaded, cleaned, combined, archived and exported to one or more platforms of presentation servers. The ultimate goal of the data preparation area is to obtain ready data to be loaded on a delivery server (OLAP engine or RDBMS).

2.7.3 The data presentation area :

The presentation area is all that the user can manipulate by the access tools. It consists of the Data Warehouse or a series of Data marts.

Data mart : Structure where many useful data are stored for the strategic management of a domain or a particular department of a company. All Data Marts form the Data Warehouse. "Data mart is defined as a logical subset of a data warehouse". "**[Kimball, 2002]** Like it munched in the **Figure 2.8**.

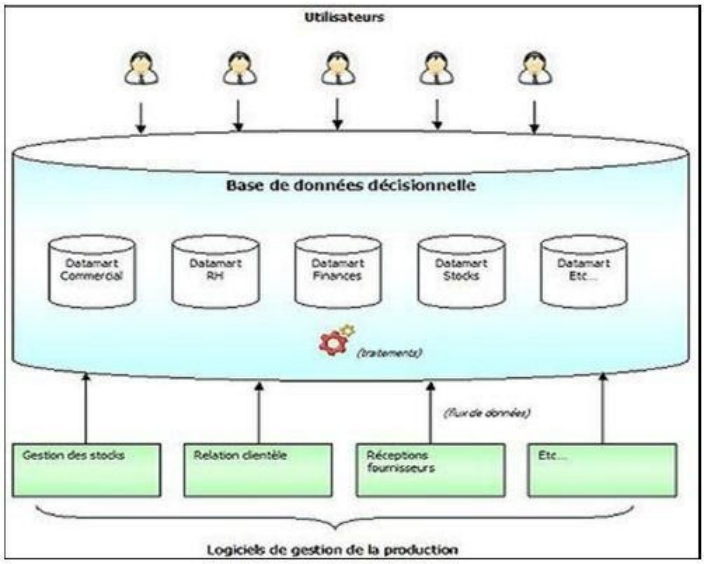


FIGURE 2.8 – Data mart in the Data Warehouse

2.7.4 Data access tools :

It's the public share of the data warehouse **[Kimball, 1996]**. We find :

— Ad hoc query tools.
— Reporting tools.

— Data Mining tools.
— Applications.

These data access tools are used by the users to access the presentation of the Data Warehouse. Their choice depends on the needs of the user in terms of information. [4]

2.8 Data Warehouse Architecture :

2.8.1 The "back room" : it allows :

— Extracting data from sources.
— Data transformation.
— Loading data into Datamarts.

The back room is the data preparation area.

2.8.2 The front room : it allows :

— Data analysis via reports or cubes (OLAP, Report, ...).
— Drilling and data mining.
— Ad-hoc queries to retrieve data in the past from the data historic.

The front room contains the data presentation area in **Figure 2.9**.

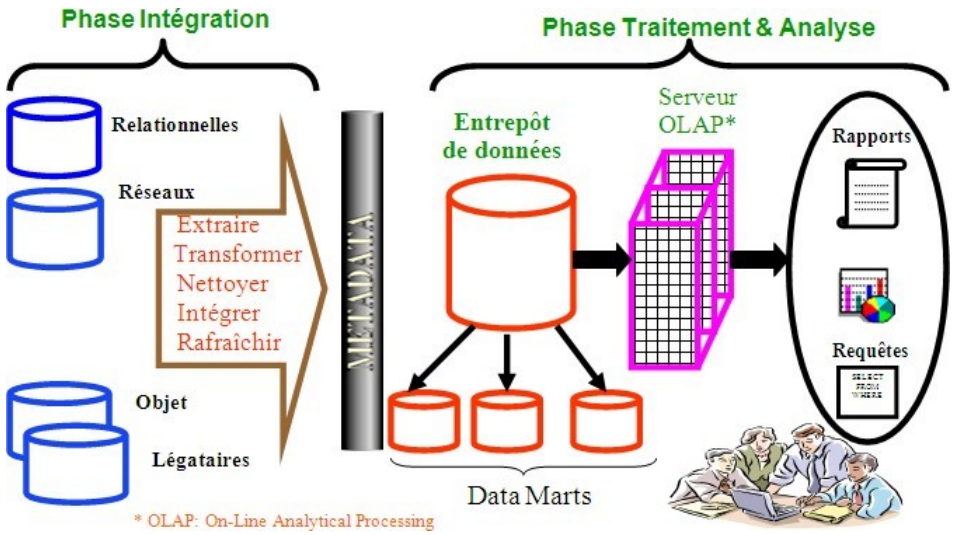


FIGURE 2.9 – Designing a Data Warehouse

4. R. Kimball ; "Data Warehousing : A Data Warehouse Designer's Guide", Wiley Computer Publishing 1996.

2.9 Data Warehouse vs. Databases :

Data Warehouses (decision-making) are different in many ways from the basics of data (Operational or transactional). Here is an overview of these differences in **Table 2.1** :

Data Warehouse	**Databases**
Oriented subject	Oriented application
Summary, refined, detailed	Detailed
evolves with time	Static
Decision-making Needs	Daily Needs
Read Only	Play, Update, Delete
Batch Processing	Real Time Transactions
Analysis	Transaction
Dimensional Model	Entity-Relationship Diagram
Large upstream of data	Small amount of data
Relative availability	High availability
Flexible structure	Static structure

Table 2.1 – Comparison between Data Warehouse and BDD

2.10 Data Modeling of the Data Warehouse :

2.10.1 Dimensional modeling :

Dimensional modeling often called OLAP, is a design method logic that aims to present the data in an intuitive standardized form and allows high performance and very fast access. It is based on two concepts fundamental facts and dimensions. This storage mode is intended for analysis statistical data.

2.10.1.1 Tables of facts and tables of dimensions :

Table of Facts : The "**fact**" represents the observation of the market on a subject analyzed. The fact table serves store the measures of the activity. A row of a table is a measure. These they are numeric,

additive values. "Measurements of a fact are digital and generally valued on an ongoing basis". [Kimball, 00].

All tables of fact are composed of at least two foreign keys, they are connect to primary keys tables of dimensions. For example **Figure 2.10**.

FIGURE 2.10 – A fact table

Table of dimensions : "The subject to be analyzed, that is to say the fact, is analyzed according to different perspectives. It's perspectives correspond to a category used to characterize the measures of activity analyzed " [Marcel, 98].

The dimension models a perspective of the analysis. A dimension consists of parameters corresponding to the information making the activity measurements different. For example **Figure 2.11**.

FIGURE 2.11 – A dimension table

2.10.2 Different models of dimensional modeling :

There are three types of dimensional models : star model, flock model, model in constellation :

2.10.2.1 Star model (Star Diagram) :

The fact table is in the schema center, and the other dimension tables are connected to the table of fact with a single join. This is the most used data structure and the easier for DW users, the Example are munched in the **Figure2.12**.

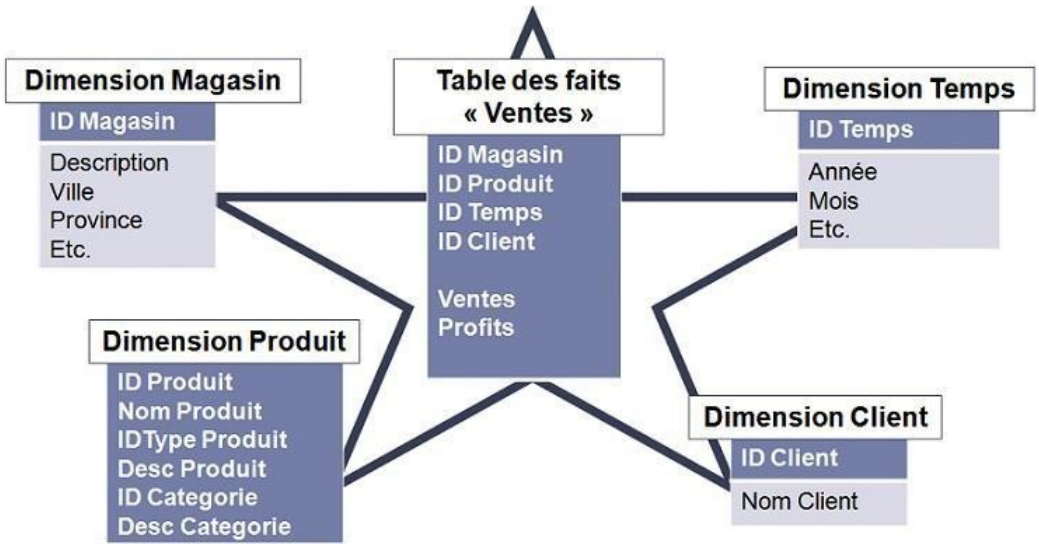


Figure 2.12 – A star schema

2.10.2.2 Flake Model (Snow Flake Pattern) :

The flake model is a variant of the star model, it simplifies the standardization tables of dimensions. I'm given the attributes of each hierarchical level in a dimension table apart as Flake diagram in the **Figure 2.13**.

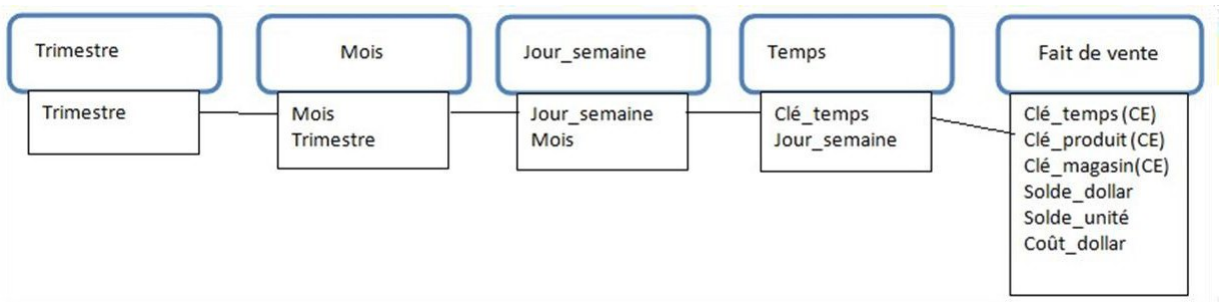


Figure 2.13 – A flake diagram

2.10.2.3 Constellation model :

The constellation model uses common dimensions for a set of models in stars.

2.11 The data cube :

The multidimensional database is based on the hyper-cube (n-dimension), this method allows analysts and decision makers to have the information they need. They selected criteria to visualize as a table or cube or hyper-cube, each side of the cube has a dimension (eg : product, region, date) and the corresponding characteristics (product : keyboard, hard disk, mouse), the graphical presentation of data cube is in the **Figure 2.14**.

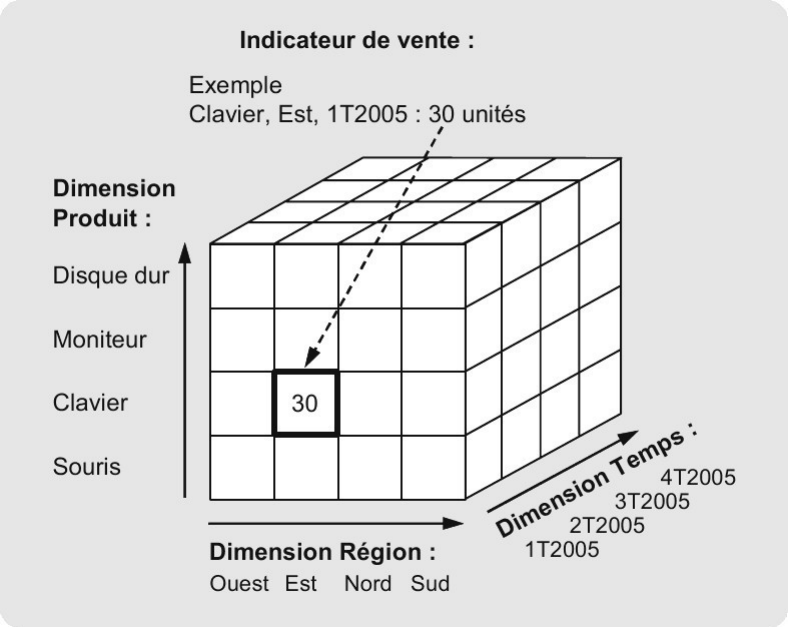


Figure 2.14 – A data cube

2.12 The OLAP concept :

The OLAP concept was introduced and defined in 1993 by EF Codd. OLAP «On Line Analytical Process "means a category of software that focuses on the exploration and rapid analysis of data in a multi- dimensional approach (multiple levels of aggregation).

The concept OLAP " is a global activity of querying and presenting data Textual and digital content in the data warehouse Style the polling specifically dimensional " **[Kimball, 2005]**.

The OLAP tool allows you to restructure and store data in a format Multidimensional.

2.12.1 OLAP storage architectures :

OLAP is doing analysis on multi-dimensional datasets. X-OLAP defines how the data will be physically stored to allow multi-dimensional analyzes. **[GRIM, 2009]**.

2.12.1.1 M-OLAP (Multidimensional OLAP) :

MOLAP is one of the first multidimensional technologies. "It's a set of application interfaces and database technologies owner whose dimensional aspect is preponderant " [Kimball, 2007]. [5].

MOLAP storage is done in dimensional databases that are used Frequently and require a minimal response time. Its disadvantage is that it does not support very large volumes of data and we justify with the principal of M-olap architecture in the **Figure 2.15**.

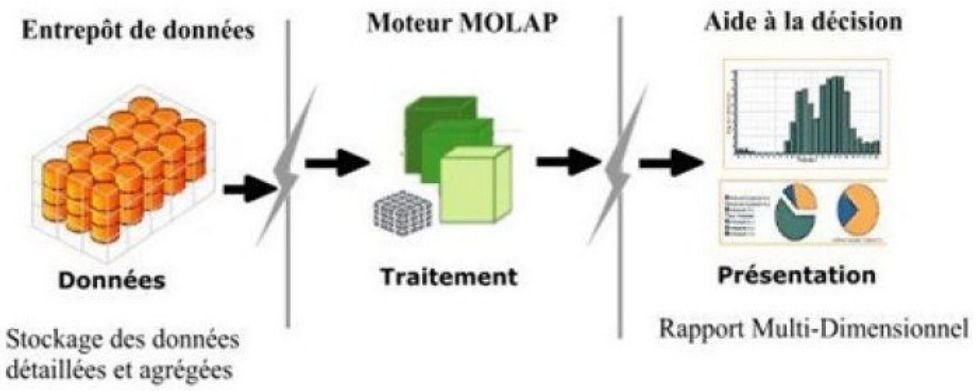


FIGURE 2.15 – Principle of the MOLAP architecture.

2.12.1.2 R-OLAP (OLAP Relational) :

The ROLAP is a set of user interfaces and applications that give a dimensional vision to relational databases [Kimball, 2005].

Aggregated Data Warehouse data is stored as tables in a relational database, it must be structured according to a particular model (star, flake, ...). Data processing is done with advanced SQL. Unlike the previous technique, the R-OLAP technique is easy to implement, and it can store large volumes of data. But she is less successful during the calculation phases munched in **Figure 2.16**.

5. R. Kimball, M. Ross and W. Thornthwaite, "The Data Warehouse - Project Management Guide", Editions Eyrolles, 2007-06-11.

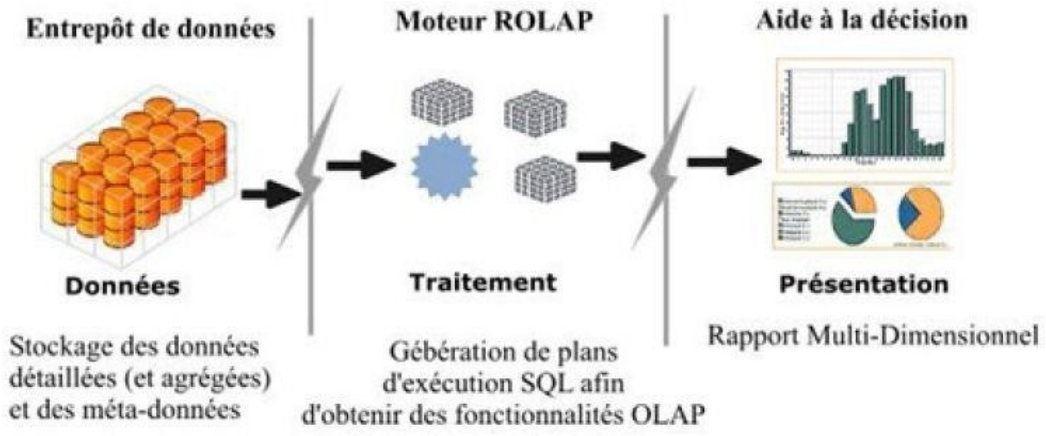


FIGURE 2.16 – Principle of the ROLAP architecture.

2.12.1.3 H-OLAP (OLAP Hybrid) :

This approach combines both relational and dimensional approaches. She combines performance in the analysis of the MOLAP database and the storage capacity of ROLAP. The basic detailed data of the Data warehouse is stored in a database of relational data and aggregated data are stored in a database multidimensional.

2.12.2 Comparison between MOLAP, ROLAP and HOLAP signed in Table 2.2 :

	MOLAP	**ROLAP**	**HOLAP**
Storage of data from based (Detailed)	Multidimensional DB	Relational DB	Relational DB
Storage of aggregations	Multidimensional DB	Relational DB	Multidimensional DB
Performance queries	High	little performing	Average

Table 2.2 – Comparative table between R-OLAP, M- OLAP and H-OLAP.

2.13 Navigation in the data :

The navigation operations in the data are :

Set operations : Slicing Dicing.

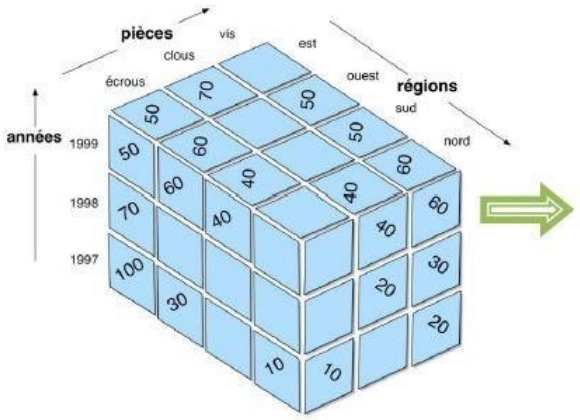
Granularity operations : Roll-up, Drill-down.

2.13.1 Set-up operations "Slice Dice" :

2.13.1.1 Slicing

involves making a selection of slices of cube according to predicates and according to a dimension 'filter a dimension according to a value ". [Chouder, 2007]

As example the **Figure 2.17**.



ventes 97-99	est	ouest	sud	nord
écrous	220	100	60	10
clous	160	50	10	60
vis	20	150	170	110

FIGURE 2.17 – Example of slicing.

2.13.1.2 Dicing

meanwhile, can be seen as an extraction of a sub cube munched in the **Figure 2.18**.

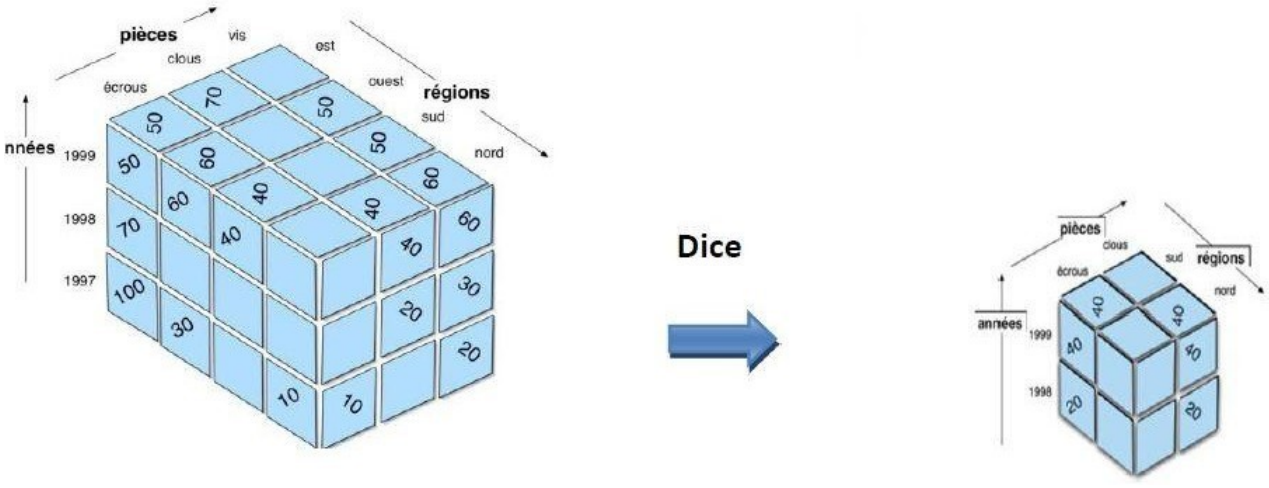


FIGURE 2.18 – Example of Dicing.

2.13.2 Roll-up Drill-down granularity operations :

These methods, also called "drilling up or down".

2.13.2.1 The Drill-down :

is to represent the data of the cube at a lower level of granularity like example **Figure 2.19**.

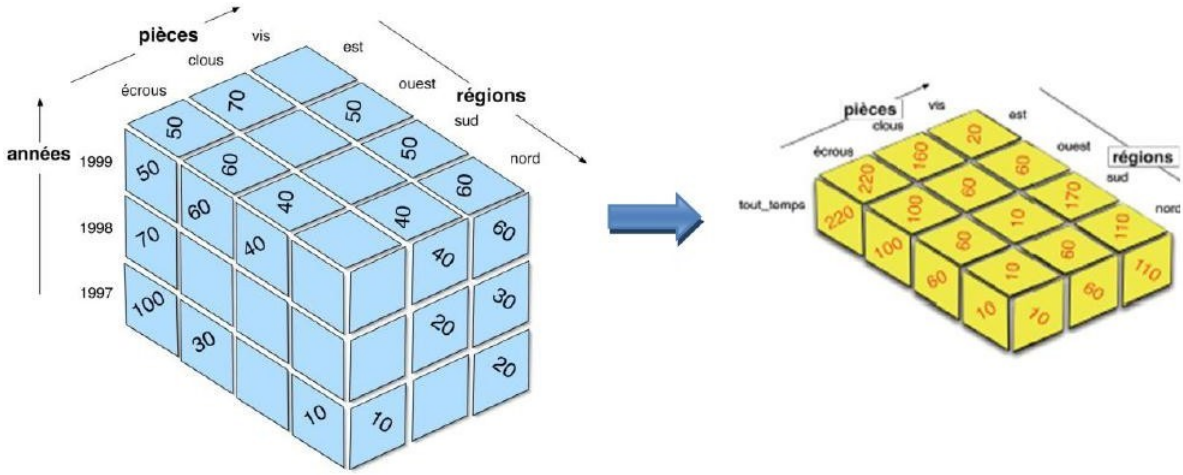


FIGURE 2.19 – Example of Drill-Down.

2.13.2.2 The Roll-up :

consists of representing the data of the cube at a higher level of granularity the great Example munched in the **Figure 2.20**.

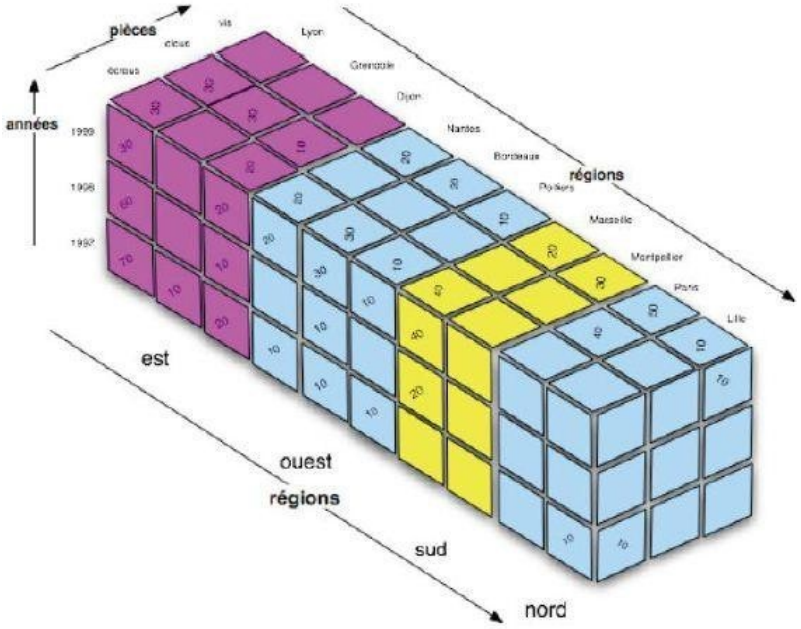


FIGURE 2.20 – Example of Roll-Up.

2.14 Conclusion :

The Data Warehouse is the heart of the decision-making system. It is subject oriented, integrated, archived and non-volatile.

In this chapter, we have discovered how to represent data in order to that they are actionable and visible to the decision-maker, and the different types of modeling to improve response times.

The dimensional modeling the proved to be the only architecture for the consistent construction of a data warehouse. Using facts and consistent dimensions of a set of templates dimensional is a safe way to build a Data Warehouse system.

Chapitre

3 Conduct and construction of the Data Warehouse

3.1 Data Warehouse Building Methodologies

There are three commonly used methods for setting up a Data Warehouse. But we must look for the best approach to the context, It is :

3.1.1 Top-Down

You can catch all the minnows in the ocean, group them together, and they do not can not make a whale. **[Bill Inmon]**

This method consists of building the schema of the data warehouse from those sources of data and assumes that the schema that will be constructed will be able to respond to all the analysis needs. It is generally supported for Data marts. The advantage that this method is that it offers a very clear and very conceptual vision company data as well as the work to be done, As example the **Figure 3.1**.

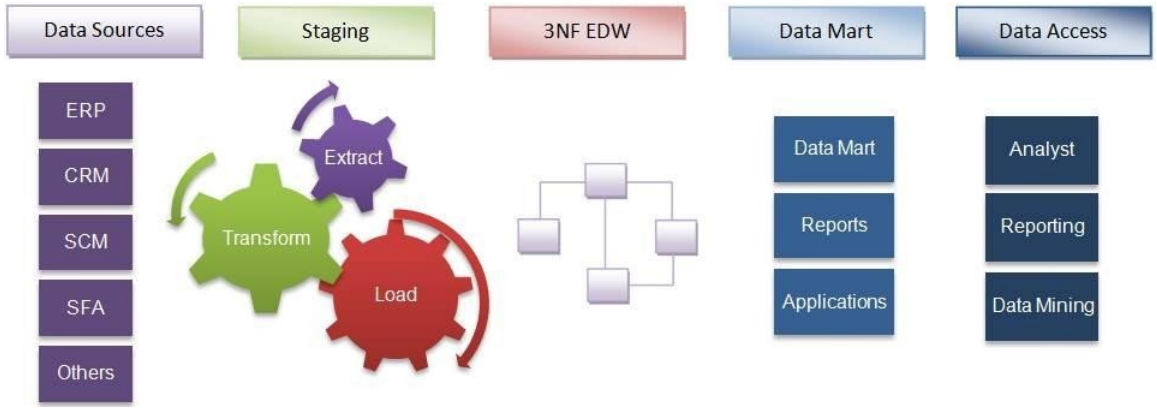


Figure 3.1 – DW architecture according to Inmon.

3.1.2 Bottom-Up

"[...] The Data warehouse is nothing but the union of all Data marts." **[Ralph Kimball]**.

This is the opposite approach of the previous one, it consists of creating the stars one by one, then to group them by intermediate levels until you get a real warehouse pyramidal with a corporate vision.

The advantage of this method is that it is simple to perform (one star at a time), we see it in the **Figure 3.2**.

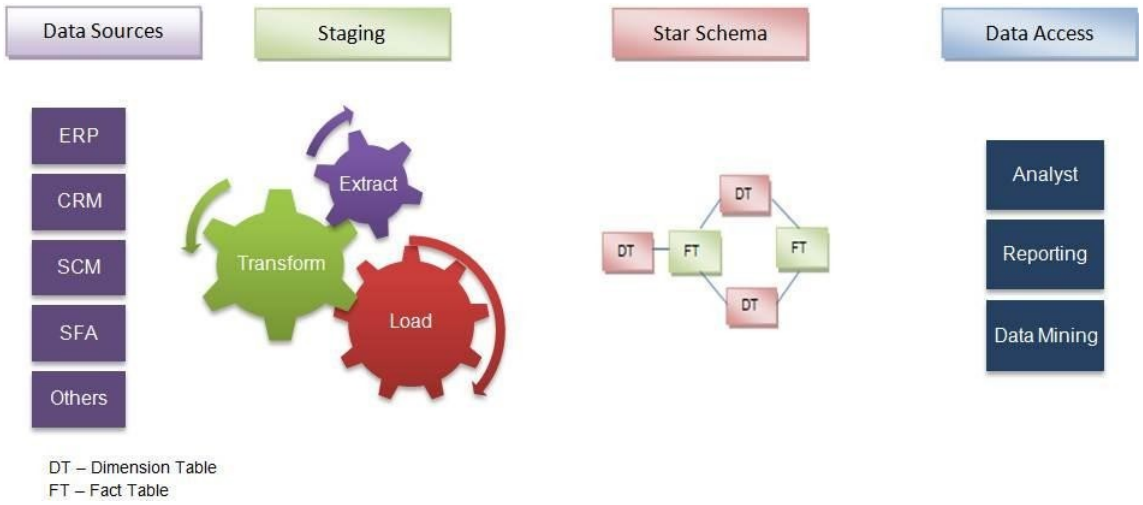


FIGURE 3.2 – Kimball's DW architecture.

3.1.3 Top-Down vs Bottom-up

Top-Down **Bill Inmon and CIF**	**Bottom-Up Ralph Kimball**
Start by designing a DW model at the enterprise level.	Start by designing a dimensional model for the data mart.
Deploy a multi-tier architecture consisting of staging area, DW, and dependent data mart.	Uses an architecture that consists of staging area and data marts.
The DW is business oriented, data marts are process oriented.	Data marts can provide a company or process view
Users can query DW and data marts.	Users can not query the staging area.

Table 3.1 – Comparative table between the approaches of setting up a Data warehouse according to Inmon and Kimball.

3.1.4 Middle-Out

This is the hybrid approach, and recommended by BI professionals. This approach hybrid considers both the analysis needs and the data for the construction of the diagram. The idea in this approach is to build candidate schemas from data. Thus, the constructed schema is a response to the real needs of analysis and it is also possible to implement it with data sources.

3.2 Data Warehouse Power Supply

The data of a **DW** come from different databases. These bases of production, operating systems correspond to all computer applications used daily in companies for its activity (production management, management banking, commercial management, ...).

The data stored there, specific to each software, can be used by other software, through data transfers, commonly called interfaces. [Houses, 2006].

3.2.1 Extract-Transform-Load ($\mathcal{ETL}$) process

These are the operations managers who are taking place in the backstage of a Data Warehouse architecture. The data are extracted from different sources of data, after they will be transformed into a more appropriate structure for analysis and reports and finally they are loaded into the database, like example the **Figure 3.3**.

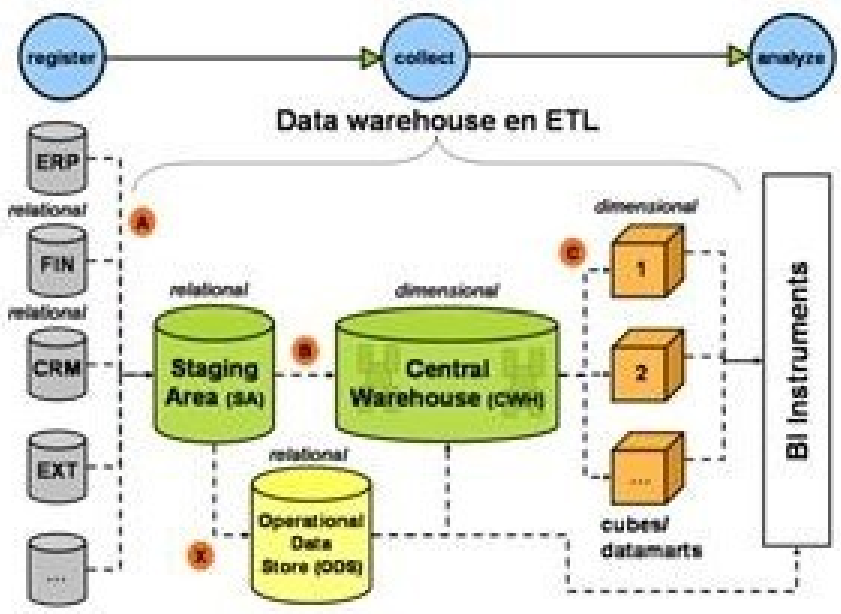


FIGURE 3.3 – The ETL process.

3.2.1.1 Data extraction

"Extraction is the first step in the process of providing data to the warehouse. Data, extracting, that means reading and interpreting the source data and copying them into the preparation zone for previous manipulations " [Kimball, 2005].

It involves extracting data from different internal and external sources, structured and / or unstructured. Several queries are sent to the in most cases with the same structure as the source.

3.2.1.2 Data transformation

It's about filtering and sorting the data available in staging using attributes specific and make calculations. In this step of the **ETL** process, you can check the quality of data and clean the data if necessary.

3.2.1.3 Loading data

This is the last step of the **ETL** process, the data is loaded into a warehouse central. We can now combine, aggregate and load into cubes or Data mart if deemed necessary.

3.2.1.4 Data Profiling and Data Quality Control

Data profiling provides a live view of data quality source systems. It can display the number of rows that have missing values or invalid. We can specify business rules to clean the data. Do profiling of data prior to the design of the **ETL** process, ensures construction a solid system with a clear structure.

3.3 The deployment and implementation of a Data warehouse

This is the last step of the Data Warehouse project. After storing, cleaning and consolidated data, they are usable. As needed, different types of tools exploitation will be considered.

3.3.1 Reporting

The term **"Reporting"** refers to a family of Business Intelligence tools intended to ensure the production, publication and dissemination of activity reports in a format predetermined. They are mainly intended to facilitate the communication of results encrypted or tracking progress,As example in the **Figure 3.4**.

FIGURE 3.4 – Example of an output status in reporting.

Reporting tools provide querying of databases according to queries **SQL** prepared during the development of the model. They offer specific functions for development of the report template, calculation and presentation modules (graphics) to develop relevant accounts.

3.3.2 Dashboard

The dashboard is a set of few indicators, designed to allow managers to take cognizance of the state and evolution of the systems they pilot and identify trends that will influence them on a horizon consistent with the nature of their functions,like example the **Figure 3.5**. [Bouquin, 2003].

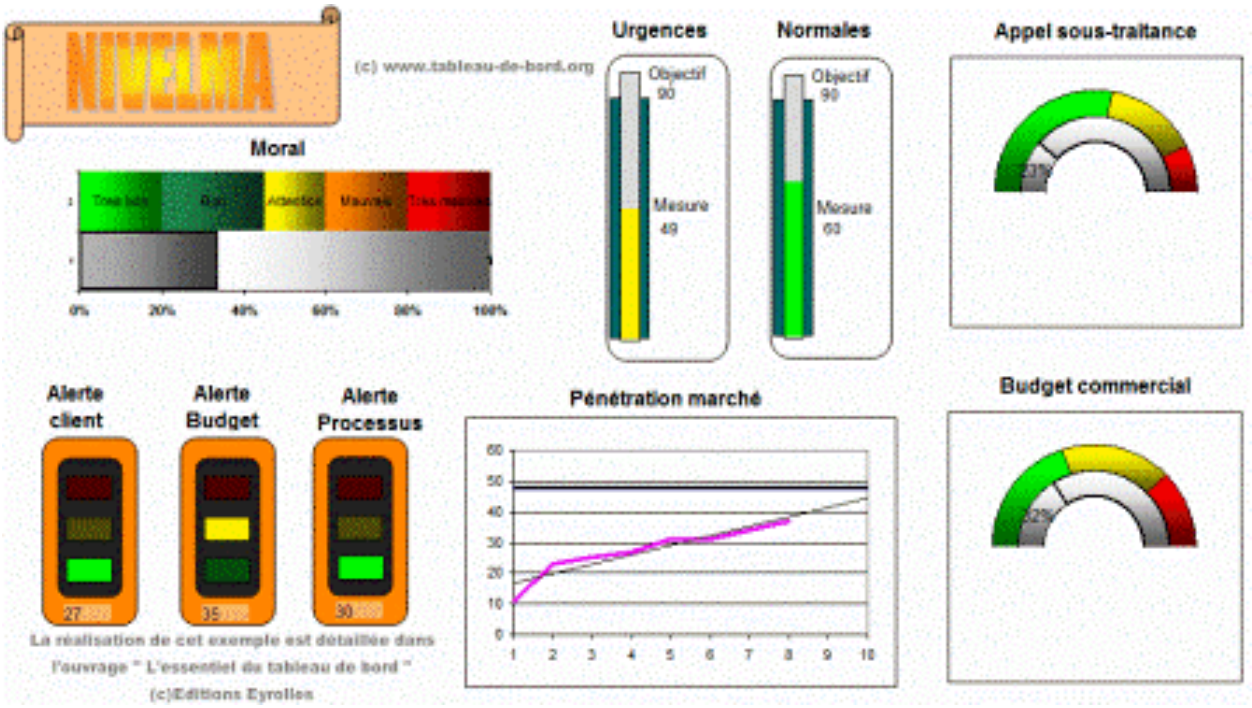


FIGURE 3.5 – Example of an output status in reporting.

The dashboards are active aid instruments the decision making and guidance, they measure all forms of performance and assist decision-makers as effectively as possible in situation.

3.3.2.1 The objectives of the dashboards

- Know everything about the situation of the company.
- Follow the achievement of objectives.
- Follow the strategic deployments.
- Adjust the lag between objectives and results.
- Manage the business and help make decisions.
- Delegate and empower employees.

3.3.2.2 Design of a dashboard

The design is done through the following steps :

3.2.2.1. Identify areas of development related to the strategy (or project company) : The axes must be specific to the company, and linked to its strategy and business. And Customized according to the skills of the leader.

3.2.2.2. Set goals :

Smart.

— Specific.
— Measurable.
— Ambitious.
— Realiste.
— In time.

3.2.2.3 Choose the most representative indicators : An indicator is information or a group of information that contributes to the decision maker's assessment of a situation.

An indicator should be clear, simple, relevant, shared and easily understandable.

3.2.2.4. Build the dashboard : In this step you have to adapt the dashboard at each hierarchical level (operational / strategic) and Integrate graphics for a fast and attractive visual , and comments.

3.4 Data Mining

Data Mining is a generic term encompassing a whole family of tools that facilitate the exploration and analysis of data stored in the Data Warehouse or Data Mart.

"Data Mining is the analysis of a set of observations that aims to find unsuspected relationships and summarize the data in a new way, so that be more understandable and useful for their holders " **[David Hand, 2001]**

Data Mining enables the analysis of the large volume information contained in data warehouses to detect relationships that it would be a priori impossible to identify without this tool.

3.5 Data mining techniques

For each datamining task, there are techniques to perform them. These techniques or methods are derived either from the field of data and statistical analysis, automatic symbolic learning or from the field of artificial intelligence. There are different datamining techniques, we present below, the techniques classified by domain :

3.5.1 Statistical techniques

the statistical field has long allowed synthetic description, using projection methods such as factorial analysis in principal components **(FCA)**, the search for dependencies through correspondence analysis, deviation measurements through the analysis of model residues and finally, the discovery of causal relationships through logistic regression. [6]

3.5.2 Learning techniques

we distinguish between :

6. https://pdfs.semanticscholar.org/4b7e/9d2b2157ce78bd6b64ba502381359be48de7.pdf

Unsupervised learning techniques, also known as "discovery of non-directed knowledge". These techniques are used to recognize the relationships expressed by the data, such as : association rules, clustering.

Supervised learning techniques, also known as "discovery of directed knowledge". They help to explain the relationships, once they are found. Such as : case based reasoning **(CBR)**, decision trees, neural networks. [7]

Among its techniques to use in datamining has :

— The associations.
— Decision trees.
— Neural networks.
— Multiple linear regression.

3.6 Definition of Multiple linear regression

Multiple linear regression makes it possible to study and measure the mathematical relationship that may exist between several quantitative variables. Multiple linear regression or the case of several explanatory variables. Multiple linear regression is a generalization, with p explanatory variables, of simple linear regression. We are still in the context of mathematical regression. We try to predict, with the greatest possible precision, the values taken by a variable y, called endogenous, from a series of explanatory variables X (x1, x2, ..., .., xp.). In the case of multiple linear regression, the endogenous and exogenous variables are all quantitative (continuous), and the prediction model is linear (x1, x2, ...,..., xp.).

3.7 Conclusion

The construction of the data warehouse is the most important phase in order to have a better implementation of all types of **decision tools**. In this chapter we have presented the different methodologies of construction of the Data Warehouse and the tools Power.

During the second part, we will introduce the host organization and the different stages of building our system.

7. https://en.wikipedia.org/wiki/Case-based_reasoning

Part II

Preliminary study

Chapitre

4 Presentation of the host company , Study of the existing system and Study of needs

Sommaire

4.1 Definition Supermarket . . . 39
4.2 Introduction . . . 40
4.3 Operating systems . . . 40
4.4 Decision Support Systems . . . 40
4.5 Computer applications and licenses . . . 42
4.6 Conclusion . . . 43
4.7 Introduction . . . 43
4.8 Presentation of information gathering approaches . . . 43
4.9 Presentation of the approaches used . . . 45
4.10 Obstacles encountered . . . 45
4.11 Needs analysis . . . 45
4.12 Identification of users . . . 47
4.13 Conclusion . . . 47

4.1 Definition Supermarket

A retail point of sale, in full self-service, on a single level, at reduced prices and margins, with a surface area of between 400 and 2,500 m2 and an assortment - 3,000 to 5,000 references on average - consisting both of food products - for the most part - and non-food products (cleaning products, hardware, stationery, lingerie, etc.). Payment is made for almost all products, in one go, at the checkouts located at the exit of the store. American Michael Cullen was credited with opening the first supermarket in 1930 in Jamaica, New York, under the name "King Cullen". The concept was taken up by Édouard Leclerc in France in 1949, but Goulet-Turpin claimed the opening of the first supermarket in France, in Rueil-Malmaison in the Paris region in 1958. [8].

Hichame supermarket presentation : supermarket Hichame is a private company, well positioned in the wilaya of Oran, with a significant annual turnover. Directed by Mr. BOUMEDIENNE Hichame, SARL Hichame stands out from its competitors by diversifying its service segments.

Suppliers :

8. https://e-marketing.fr/definitions-glossaire/supermarchés

Supermarket Hichame to several suppliers throughout the territory of Algeria.

we have chosen this supermarket for the collection of information and data necessary for the realization of our work which is usable for any company that does commercial management and inventory management.

4.2 Introduction

In this chapter we will present the operational systems deferents. And decision-making procedures. In order to understand the current processes of the company's commercial management and to analyze the various decision-making processes. Our study was limited to the scope of the project, i.e. the Sales Department for which the solution is intended.

4.3 Operating systems

The company **(Supermarket Hichame)** has a single operational system, SAGE, which manages all the company's business processes. It is composed of 3 essential modules :

- Wise accounting : allows you to record their daily accounting operations (record receipts, credit sales and disbursements).
- Wise commercial management : allows the commercial management of the purchase order to the posted supplier invoice and the estimate to the invoiced customer. While including different control tools such as stocks and stock level.
- Sage Payroll : allows the management of information relating to employees and ensures the management of payroll and the administrative management of personnel (civil status, salary, leave, social and tax data, social balance sheet, certificates, etc.).

4.4 Decision Support Systems

The company has no decision support system, which makes it impossible to manage the large amounts of information circulating in its production system, to structure them, and to generate new knowledge to make them available to users.

Decision makers use the predefined reports extracted from the operational system to analyze the company's situation as shown in **The Figure 4.1 and The Figure 4.2**

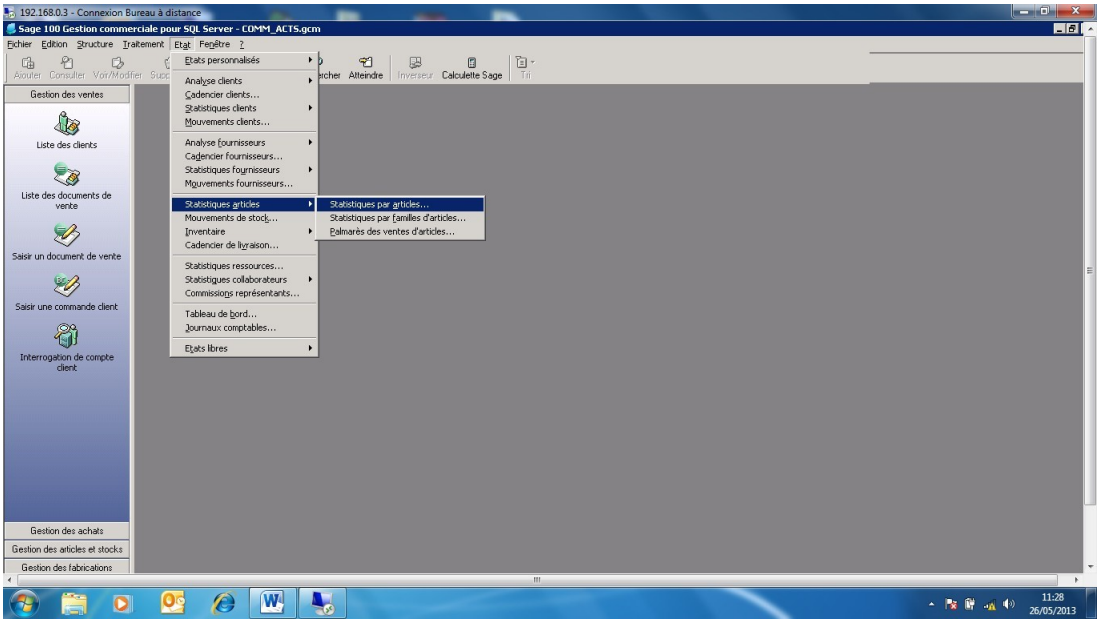


Figure 4.1 – Use of a SAGE report to obtain turnover.

FIGURE 4.2 – Choice of valuation and period.

4.5 Computer applications and licenses

The main applications used are :

- SAGE.
- Exchange - Outlook.
- Windows operating system.

4.6 Conclusion

In this chapter we have presented in detail the different operational systems that manage all the company's business processes, and the state of decision-making within the company.

This study allowed us to have a vision of the current systems. It is the starting point of our project.

Next we will present the needs concerning the decision-making aspect of the company and the approaches used for data collection.

4.7 Introduction

"The chances of success of a Data Warehouse are greatly increased by a good understanding of users and their needs"[Kimball, 2007]. The collection and identification of needs phase is essential for the construction of our system. In order to meet the expectations of decision-makers.

4.8 Presentation of information gathering approaches

"Organized process implemented to obtain information from multiple sources, with a view to moving from one level of knowledge or representation of a given situation to another level of knowledge or representation of the same situation, as part of a deliberate action whose objectives have been clearly defined, and which provides sufficient guarantees of validity".

Data collection is vital to have a detailed understanding of the company's situation and to have the necessary information for project design.

It also makes it possible to know what works and what does not.

The study of documents : This method is essential to obtain theoretical and practical data. It consists in studying the reports used by the various departments of the company in order to know how it works.

Interviews : Interviewing is one of the most common techniques used to collect qualitative information. The interview is a formal conversation between an interviewer and the selected person, oriented towards a goal. In order to be able to feed the search for information on a fixed theme.

Advantage :

- It does not require a lot of resources or staff.
- It allows for more personal contact with the target populations.
- It provides very detailed and detailed information.

Disadvantage :

- It is a difficult technique to prepare and implement.
- It often requires time.

Observation : It is a qualitative method of collecting observable data such as gestures, spaces, sentences or interjections (the forms of interaction between caregivers and (the number of

people in a waiting room), temporalities (average waiting time for a service), etc. It is based on the careful monitoring of people's facts and practices, without any desire to change them. Two categories can be distinguished :

Directed observation : Method is based on direct observation in the field of the facts being studied.

Participant observation : This method implies that one does not simply observe one's terrain by remaining outside the scope of the project, but that one must participate in ongoing activities on the ground by adopting a role that already exists in the situation under study.

Advantage :

— It produces information on actual practices.
— It shows discrepancies between saying and doing.
— It does not require many resources.

Disadvantage :

— there is a risk of missing important data or directing results.
— It often requires time.
— Difficulties in interpreting certain behaviours.

The questionnaire : The questionnaire is a tool for quantifying and comparing information. This information is collected from a representative sample of the population targeted by the evaluation. It is a set of questions constructed to obtain information corresponding to the evaluation questions. There are two forms of questionnaire :

The closed questionnaire : It requires precise answers and a limited number of answer choices.

The open questionnaire : In this category people develop answers and the interviewer takes note of them. An open-ended question leaves the answer free in its form.

Advantage :

— Allows to collect a large amount of information.
— Useful for measuring change and making comparisons between opinions.
— Allows a large number of final beneficiaries to be consulted.

Disadvantage :

— Requires significant resources.
— The wording can influence the answers.
— The answers are neither complete nor detailed.

The meetings : They consist of planning meetings in large groups, led by a facilitator. These meetings can promote free and creative expression with a limited number of participants.

4.9 Presentation of the approaches used

By studying the different approaches to information collection, we used the interview approach that allowed us to establish personal contact with the targeted users. Complemented by documentation that allows for the collection of complete information and observations due to the difficulty of interviewing some people.

First approach "Documentation" : We used this approach when studying the company's operating system. And since the company operates under SAGE, the documentation was available.

Second approach "the interview" : Because of its advantages mentioned above. And by meeting with user representatives, we were able to collect the detailed information. During our investigation, we were able to interview :

- Database administrator.
- The manager.
- Procurement manager.

The stages of the interview : In this step we tried to anticipate language problems and present and explain the reasons for the investigation.

Adjustment phase between the subject and the responses of the target persons : We tried to adapt to the status of the interlocutor and to ensure the specificities of the answers and to be flexible, while controlling the evolution of the interview. Follow-up phase of the interview and in-depth guide :

In this step we made direct observations, including those not covered by the interview guide. And show reactivity by using contradiction and relaunch.

Conclusion phase : In this step we had the respondent validate the account- interview report, quickly review and format the notes taken during the interview.

4.10 Obstacles encountered

- The busy schedules of the interviewees.
- Unforeseen events, such as travel, missions, leave.
- The unavailability of people affected by interviews and cancellations.

4.11 Needs analysis

The needs study carried out enabled us to identify the following needs :

Sales Analysis :

Detailed turnover : Know the detailed turnover by product and by product family. In addition, it is a question of knowing :

- Detailed turnover by product.
- Detailed turnover by product family.

Sales volume : Establishing sales statistics by product and by product family makes it possible to measure the growth and evolution of the company's activity over a given period. In addition, it is a question of knowing :

- Sales volume by product.
- Sales volume per product family.

Sales margin :

The trading margin is a simple indicator. It makes it possible to position a company in relation to its competitors and to determine the price at which the company must sell its products.

Margin rate : The margin rate allows the company to know the profitability of a product or a family of products. In order to know the products and product families that bring him the most benefits.

Supply analysis :

Number of units imported :

Establishing purchasing statistics by product and by product family makes it possible to know and measure the growth and evolution of the company's purchases during a given period. In addition, it is necessary to know the number of units imported per product.

Standard cost per product :

The standard cost is the cost of acquiring the product before it is sold. Good purchasing management allows the company to significantly reduce its costs and make the company more competitive.

Stock Analysis :

Volumes of products available in stock :

Establishing statistics on the volume of available product allows you to have a vision on the state of stocks, i.e. its level, in order to optimize its management and control problems related to stock shortages.

Stock rotation : Stock rotation is one of the indicators used by management control functions. It corresponds to the average frequency of stock renewal over a given period. It reflects the overall efficiency of the supply chain and the quality of supply management.

Replenishment alert :

Replenishment alerts facilitate stock management. They are set on the quantity for each item. When the quantity of offers that are ready to be shipped reaches the parameter threshold, they inform us by sending a message. The threshold is based on average sales and the time required to replenish stock for this article.

Summary of the needs detected :

Activity	Needs expressed
Sales	- Detailed sales by product and by product family. - Sales volume - Sales margin. - Margin rate.
Supply	-Number of units imported. - Standard purchase cost
Stock	-Volumes of products available in stock -Stock rotation. - Replenishment alert

Table 4.1 – Summary table of needs detected.

4.12 Identification of users

Through this study, we were able to detect the future users of our new system. These users are exclusively employees of the company. We can distinguish four categories :

- Owner.
- Manager.
- Procurement manager.
- Stock manager.

4.13 Conclusion

Through this chapter we were able to collect all the Information through meetings and questionnaires that allowed us to identify the users of our future system and their needs. The next phase is largely based on this preliminary study. It describes the necessary steps for the implementation and implementation of our system.

Part III

Design and Implementation of the project

Chapitre

5

Design of the zoned storage

Sommaire

5.1 Introduction 51
5.2 Participating dimensions : 52
5.3 Process modeling 52
5.3.1 Modelization of process " Sale " 52
5.3.2 Modeling of the "Purchasing" process 57
5.3.3 Modeling of process " Stock " 59
5.4 Conclusion 61

5.1 Introduction

The multidimensional modeling introduced by Ralph Kimball consists of two new concepts know the facts and the dimensions. Each model is compound of a table containing a key, the table of the facts who allows of measure the activity and a set of tables that contain the contextual information varying the measures of activity in question. For well succeed the model dimensional he is necessary of achieve the four following steps :

To choose the Process at modeling : At first You have to choose the business process to study. A process is a series of natural activities carried out in the enterprise, using different resources human, hardware and financial. The choice of processis done by users end.

To define the granularity of process : In this step he must reply at the question : that represent a recording of the fact table ?

It is step the more critical then of the creation of model. She defines the level of detail contained in the table of made.

Choose Dimensions : In this step one must choose the different dimensions that represent the context in which the fact took place.

Identify the Facts :

5.2 Participating dimensions :

Before of to start the design of the process, we have elaborate a board which represent all the dimensions participating the different shutters of the company. The purpose being to avoid a potential repetition of the dimensions :

DIMENSION/PROCESSUS	VENTE	ACHAT	STOCK
TEMPS	X	X	X
PRODUIT	X	X	X
DEPOT	X	-	X
CILENT	X	-	-
FOURNISSEUR	-	X	-
DELEGUE	-	X	-
MODE - PAYMENT	-	X	-
COMMANDE	-	X	-
PRODUIT- RETOUR	-	X	-
POINT-VENTE	X	-	-
TYPE-VENTE	X	-	-
FACTURE	X	-	-
MODE-VERSEMENT	X	-	-
PERIODE	X	-	-
UNITE	-	-	X
TYPE - STOCK	-	-	X

Table 5.1 – The dimensions participating in the different components.

5.3 Process modeling

5.3.1 Modelization of process " Sale "

Definition of the process : " Action of to sell some thing, to change a commodity against of money. "

[dictionary Larousse.]

Activity Grain : To follow the achievement and evolution of Figure business in amount and in value, by customer, by point of sale, and by product while a period given.

Participating dimensions : We mention the different dimensions describing the made and who we will to clearly define the process "Sales ".

Dimension «Time» : " The alone dimension who figure systematically in all warehouse of data, because in practice all warehouse of data is a series time. The time is the more often the first dimension in the underlying ranking of the database " **[Kimball, 2000.]**

The following table contains the details of the "Time" dimension :

Temps	
Désignation	**Détail**
N_temp	Identifiant de la dimension « Temps
date	La date exacte du jour concerné
jour	Nom du jour
AnneeDate	La date du l'année
AnneeNom	Le nom du l'année
SemestreDate	Le Numéro du semestre
SemestreNom	Le Nom du semestre
TrimestreDate	Le Numéro du trimestre
TrimestreNom	Le Nom du trimestre
MoisDate	Le Numéro du mois
MoisNom	Le Nom du mois
SemaineDate	Le Numéro du semaine
SemaineNom	Le Nom de la semaine
JourDelannee	Le Numéro de jour de l'année
JourDelanneeNom	Le Nom du jour de l'année
jourDesemestre	Le Numéro de jour du semestre
jourDesemestreNom	Le Nom du jour du semestre
jourDuTrimestre	Le Numéro de jour du trimestre
jourDuTrimestreNom	Le Nom du jour du trimestre
jourDuMois	Le Numéro de jour du mois
jourDuMoisNom	Le Nom du jour du mois
jourDeLaSemaine	Le Numéro de jour de la semaine
jourDeLaSemaineNom	Le Nom du jour de la semaine
SemaineDelannee	Le Numéro de la semaine de l'année
SemaineDelanneeNom	Le Nom du la semaine de l'année
MoisDelannee	Le Numéro mois de l'année
MoisDelanneNom	Le Nom du mois de l'année
MoisDuSemestre	Le Numéro du mois du semestre
MoisDuSemestreNom	Le Nom du mois du semestre
MoisDuTrimestre	Le Numéro du mois du trimestre
MoisDuTrimestreNom	Le Nom du mois du trimestre
TrimestreDelannee	Le Numéro du trimestre de l'année
TrimestreDelanneeNom	Le Nom du trimestre de l'année
TrimestreDuSemestre	Le Numéro du trimestre du semestre
TrimestreDuSemestreNom	Le Nom du trimestre du semestre
SemestreDelannee	Le Numéro du semestre de l'année
SemestreDelanneeNom	Le Nom du semestre de l'année
AnneeCode	Le code de l'année
SemestreCode	Le code du semestre
TrimestreCode	Le code du trimestre
MoisCode	Le code du mois
SemaineCode	Le code du semaine

Table 5.2 – Descriptive table of the "Date" dimension.

"Product" dimension : " The dimension product is Moon of the two or three dimensions of virtually any info market "

[KIMBALL 02].

The following table contains the details of the "Product" dimension :

Produit	
Désignation	**Détail**
N-Prod	Référence de Produit
Désignation	Code barre du Produit
Catégorie	Nom du Produit
Date-Expiration	Date d'expiration du produit
Qte-Tot	Quantité totale du produit
Qte-Vent	Quantité vendez
Qte-Rest	Quantité restez
HTT	Prix hors taxe
TTC	Prix avec taxe
Bénéfice	Prix bénéficié

Table 5.3 – Descriptive table of the "Product" dimension.

Dimension "Deposit" : This dimension contains all the news concerning the Showroom, and the deposits. The following table contains the details of the "Deposit" dimension :

Dépôt	
Désignation	**Détail**
N_dépot	Numéro interne du dépôt
Intitule	Intitulé du dépôt
Adresse	Adresse du dépôt
Commune	Commune
Ville	VILLE

Table 5.4 – Descriptive table of the "Deposit" dimension.

"Customer" dimension : As the customer is an element essential in our analysis. Because the officials need analyze evolution of their customer by report the services and products proposed. This dimension contains all the news concerning the customers physical and moral.

The following table contains the details of the "Customer" dimension :

Client	
Désignation	**Détail**
N_Client	Numéro du Client
N_Tel	Numéro de tel du client
Nom	Nom du client
Prénom	Prénom du client
Adresse_Client	Adresse du client

Table 5.5 – Descriptive table of the "Customer" dimension.

"Point of sale" dimension :
The point of sale is also an element essential in our analysis. Because the officials need to analyze the income in relation to the geographical areas. This dimension contains all the news concerning these points of sales.
The following table contains details of the "Point of Sale" dimension :

Point de vente	
Désignation	**Détail**
N_Point_vente	Numéro du Point de vente
Nom	Nom du point de vente
Adresse_point	Adresse du point de vente
Commune _point	Commune du point de vente
Ville_point	Ville du point de vente

Table 5.6 – Descriptive table of the "Point of sale" dimension.

Dimension "Invoice" :
As the invoice is an element essential in our analysis. Because the officials need to keep traceability and accounting documents.

The following table contains the details of the "Invoice" dimension :

Facture	
Désignation	**Détail**
Id_com	Numéro du Facture
Etat	Etat de la facture (servies/non servies)

Table 5.7 – Descriptive table of the "Invoice" dimension.

The measurables : Based on our needs study, we have detected the following measurables :
- " Qty " matches the number of product.
- " Ttc " matches the price of total sale.
- " Bénéfice " matches the profit retained by product.

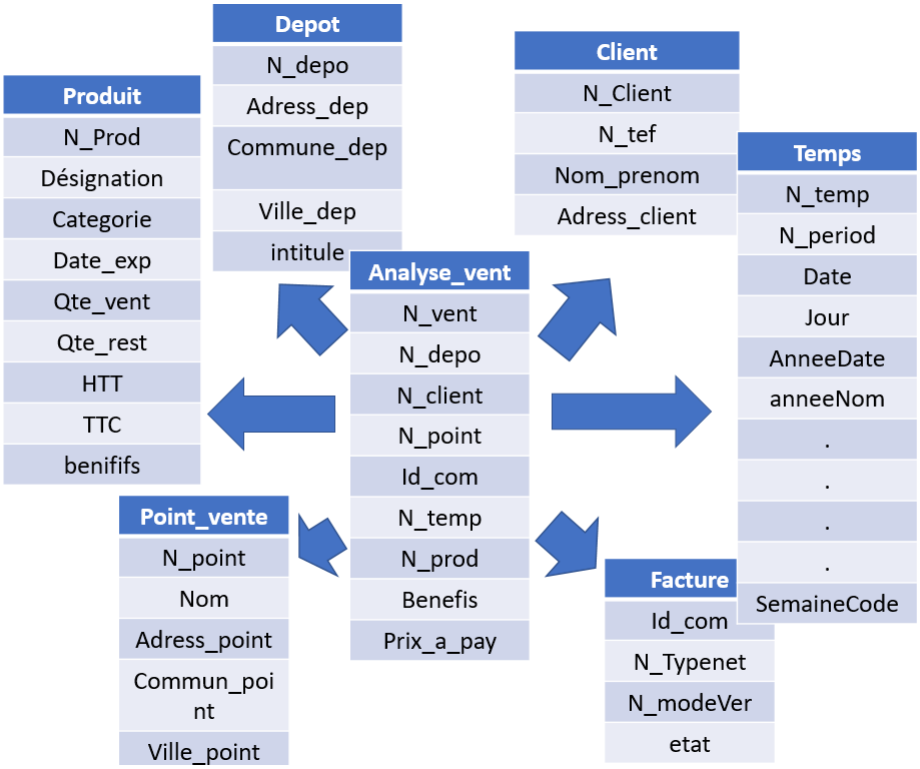


FIGURE 5.1 – Star model of the "Sale" process.

The aggregates :

Dimension	Aggregate
Temps	N_temps
Client	N_client
Dépôt	N_depot
produit	N_produit
Point_vente	N_point_vente
Facture	Id_com

Table 5.8 – Listing of the aggregates of"Sale".

5.3.2 Modeling of the "Purchasing" process

Definition of the process : " It is the act by which a nobody physical or morals gets a well or a service against the payment of a sum of money. "

Activity Grain :

ensure of the availability of the products in reply the requirements customers and the turnover stopped by Provider and by family of product while a period given.

Participating dimensions :

In more of the dimensions " Time " " product " already views the course of modeling the process "Sales" we have :

The dimension " Provider " :

Fournisseur	
Désignation	**Détail**
Id_four	Numéro id du Fournisseur
N_delege	Numéro id du délégué
Nom_prenom_four	Nom et prénom du fournisseur
Ville	Ville de fournisseur

Table 5.9 – Descriptive table of the "Provider" dimension.

The dimension " order " :

Commande	
Désignation	**Détail**
Id_Com	Numéro id du commande
Id_prod_ret	Numéro id du produit retour
Id_payment	Numéro id du payement
Etat	Etat du commande

Table 5.10 – Descriptive table of the "Order" dimension.

The measurables : Based on our needs study, we have detected the following measurables :

- " Mantant_a_payer" matches to the total amount of the order.
- " Montant_verser " matches the amount paid from each order.
- " Montant_rest " matches the amount remains to be paid.
- " $\mathrm{Nb}_{f}actIsthenumberoforders.$"$prix\_achat$"$matchesthepurchasepriceofeachproduct$.

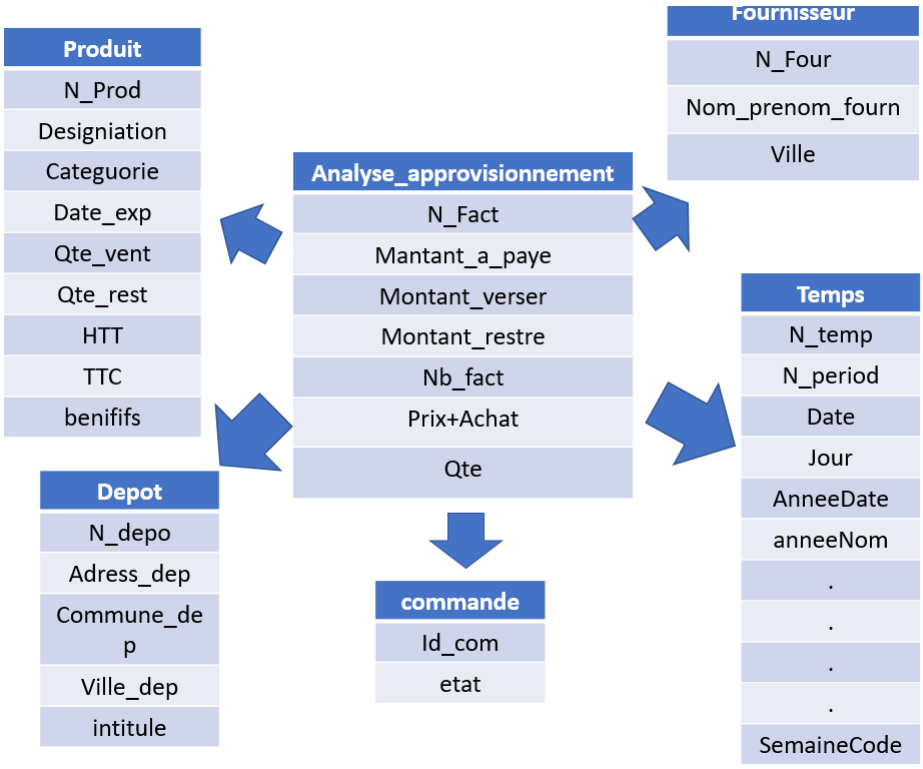


FIGURE 5.2 – Star model of the "Buy" process.

Dimension	Aggregate
Time	N_temps
Provider	N_four
order	Id_com
Product	N_produit

Table 5.11 – List of "Buy" aggregates.

5.3.3 Modeling of process " Stock "

— **Activity Grain :**

Adjust the levels of stock the needs customers and the requirements economic of the company.

Participating dimensions :

" Time " " Deposit " " product " already views the course of the modelization the development process of the " Sale " " Purchase " , We have :

The dimension " type$_S$*tock*" :

Fournisseur	
Désignation	**Détail**
N_Tstock	Numéro id du type de stock
CatégorieTstock	C'est la catégorie du stock

Table 5.12 – Descriptive table of the "type of stock" dimension.

The measurables :

Based on our needs study, we have detected the following measurables :

- "Qte_TotalSt" is the total Qty of the Stock.
- "Ttc_ST" the price of the total sale of the products.
- "Htt_ST" the total purchase price of products.

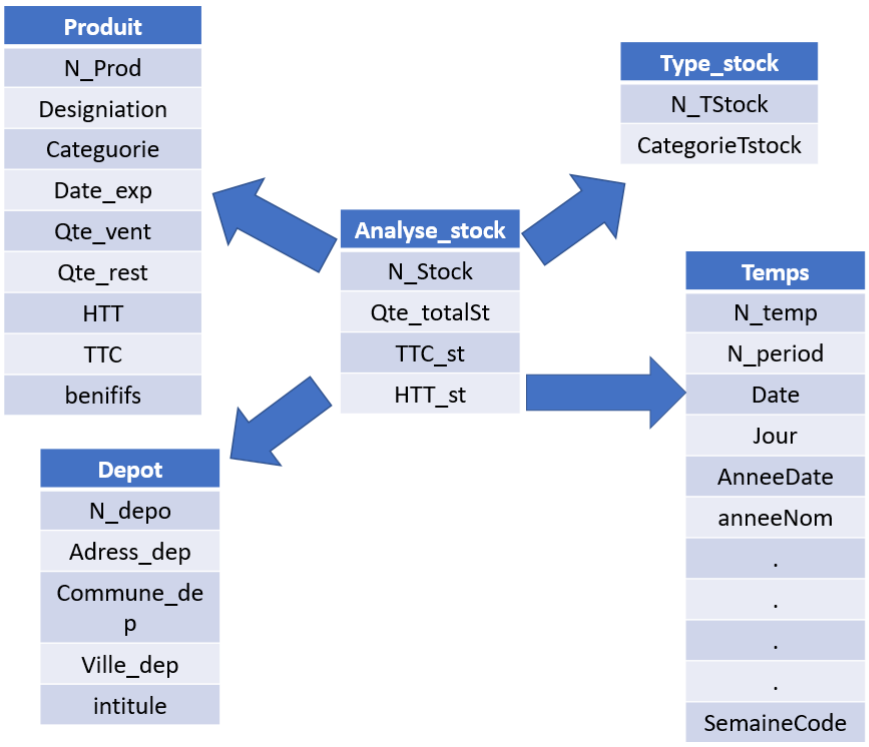


FIGURE 5.3 – Star model of the "Stock" process.

The aggregates :

Dimension	Agrégat
Temps	N_temp
Dépôt	N_depot
Produit	N_produit
Type_stock	N_TStock

Table 5.13 – List of aggregates of the "Stock".

5.4 Conclusion

In this chapter we have identified all the dimensions, facts and measurements of the warehouse of data in taking account of the process sale, purchase and stock. We have as well built the diagrams multidimensional in star who is a very model performance. it is of the made it there at less of joins to do . This phase is decisive of project of setting in square a warehouse of data. It allowed us to define the axes of analysis and the measured indicators. In order to better meet the needs of users. In the next chapter we will to describe the process power and identify the sources of data.

Chapitre

6 Design of the zoned Power

Sommaire

6.1 Introduction 63
6.2 Studies of the sources of data 63
6.3 Data Warehouse power supply 64
6.4 Conclusion 65

6.1 Introduction

The Data Warehouse's power supply is an essential phase in the construction of a decision-making system, accounting for **80 per cent** of the workload. ETL processes resume the data of all the systems operational and the pretreatment for the tools for analysis and reporting. For this we went through three stages :

- Studies of the sources of data.
- Define a Data Architecture Warehouse.
- Data Power Warehouse.

6.2 Studies of the sources of data

It is step who take more of time, because he must to study and analyze the bases of data. For we learn with the structure of the bases of data, and to define all the tables that the up, their utilities, and the different associations who the connections.

"Identify likely sources of candidate data that will support the necessary decisions at the community business. he must Identify among the sources of data, elements specific than you think be the center of the data user final " **[KIMBALL .04]**

Our sources of data are the bases of data of ERP WISE than now integrated .

We there have access at through MySQL server. **We** have meet some problem who himself summarize in :

- The detection of the tables who contain the information who will be the sources of data of the marts was a task arduous, at cause of great number of tables and views.

— Access direct the sources not allowed at cause of the confidentiality of the information.

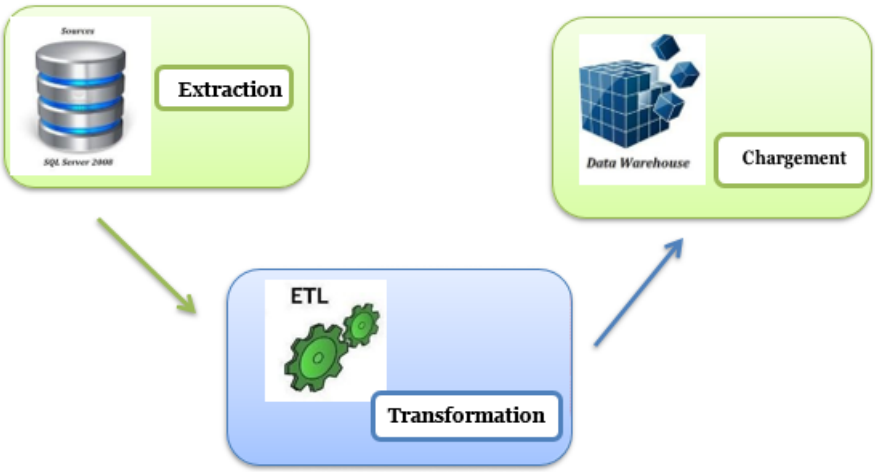


FIGURE 6.1 – Architecture of the power solution.

6.3 Data Warehouse power supply

Use a warehouse of given involved extraction a great amount of data so than the transformation of his data before their integration.

Extraction of data :

After to have designed the diagram of data warehouse, we have collection and listing the useful data in the system operational to of fill the tables of warehouse (tables of facts and the table of dimensions). This step is paramount given the amount of important data that we are going treat.

Data transformation :

" The transformation is the primary step or the system ETL added of the value. This is the step that actually changes the data and provides guidance if data can be used the purposes provided. " **[Kimball .04]**.

In this phase we will filter and sort the data available in the staging using of the attributes specific and make of the calculations. In this step we can check the quality of the data and clean the data if necessary.

Loading data :

It is the latest step of process **ETL**, the data are loaded in a central warehouse. We can now combine, aggregate and load in cubes or Datamart if this is judged necessary.

In this step we performed the following types of loading :

Loading of the dimension time :

" The dimension time the more town and useful is that of the dimension time calendar with a granularity daily. " **[KIMBALL .04]**.

This dimension does not depend on our ERP or other external systems, but the to build go of calendar. he born occurs a alone time in the cycle of life a Data Warehouse.

Loading of the tables dimensions :

" The tables of made support the measures a business. The relationship enter the fact tables and measurements are extremely simple. If a measure exists, it can be modulated as a line of the table of made. Yes a line table of made exist, so it is a measurement. " **[KIMBALL .04]**.

We will the load at go of the zoned of preparation of data who contains data from the system of production.

Loading of the tables of facts :

" The tables of made support the measures a business. The relationship enter the fact tables and measurements are extremely simple. If a measure exists, it can be modulated as a line of the table of made. Yes a line table of made exist, so it is a measurement. "

The fact table is used to store the measures of the activity. A row of a table matches at a measured. That is to say the Keywords of the dimensions participating the made will constitute the indicators of the tables of facts.

Table of made of transaction :

All of our warehouse rests sure the tables of facts of transaction. We must define the Keywords of the dimensions of each of the transactions operational and extract the measures that we interested.

Table of fact periodic : A periodic fact table is usually constructed from a transaction fact table. She represent the image a table of made of transaction at a moment T ; or the synthesis of a transaction fact table through the aggregation of measurements on its dimensions. She allows analyze a volume of transaction a lot more important and thus to release trends.

The periodic fact table must not be subjected to update operations, the production of this type of table consists in aggregating and loading the data.

Figure 6.2 – Descriptive diagram of the DW feeding process.

6.4 Conclusion

In this chapter we have described the process power of Data Warehouse and U.S have identified the sources of data. In ensuring the quality, integrity of the data and the good routing of the data towards

the Data Warehouse. In the next chapter we will to put in square the design of the cubes to enable better exploitation of data.

Chapitre

7

Design of the cubes dimensional

Sommaire

7.1 Introduction 67
7.2 levels and hierarchies of the dimensions 67
7.3 The listing of the cubes 68
7.4 Presentation of the cubes dimensional 69
7.5 Conclusion 71

7.1 Introduction

To of better navigate in the data and exploit of way effective the information contained in the Data Waterhouse, it will be necessary to set up the design of multidimensional cubes well structured and well defined for that we have opted for the model in star who present a simplicity and performance and it is of the made it there has fewer joins to make. The design of the cubes dimensional past by the definition of the measurable, dimensions and of the hierarchies present the breast of the dimensions, so than the different levels of detail of each hierarchy. She is based mainly, sure the needs identified nearby of the users and decision-makers the company.

7.2 levels and hierarchies of the dimensions

" exploitation of the cubes must to permit analyze the data a level of high detail towards a level of detail more end. To of to define these different levels each dimension is structured according to a hierarchy (sometimes several for the same dimension). the hierarchy serves at increase or at decrease the level of detail of analysis. More we mounted in the levels hierarchical of the dimensions and more he will achieve aggregation of the measurements values of each fact " [Tests .2000].

The definition of the hierarchies and levels a dimension is a step very important in the design of the cubes. In effect, it is grace these hierarchies than the user can navigate and to explore the news of cube with the level of details wish.

A hierarchy himself compound of many levels in consequent we will to define, the different levels and their attributes, then we spend the hierarchies.

A dimension is a together of values decomposable. The values a dimension are usually organized at

interior a hierarchy. access the level higher in a hierarchy is called " Rolling up " and the level inferior " drill down ".

Dimension	Attribute	Level	Hierarchy
« temps»	N_temp	Level 1	Hierarchy 1 : H1 = N1 à N2 à N3 à N4 à N5 .
	Date		
	Jour		
	anneeDate	Level 2	
	SemestreDate		
	trimestreDate	Level 3	
	MoisDate	Level 4	
	semaineDate	Level 5	
« produit»	N_produit	Level 1	Hierarchy 1 : H1 = N1 à N2 à N3 .
	Désignation		
	Catégorie		
	Date_exp	Level 2	
	Qte_vent		
	Qte_rest	Level 3	
«Dépôt »	N_depo	Level 1	Hierarchy 1 : H1 = N1 à N2 .
	Intitule		
	Adresse		
	Ville	Level 2	
	Commune_dep		
«Client »	N_Client	Level 1	Hierarchy 1 : H1 = N1 à N2 à N3 .
	Nom_prenom		
	Adresse	Level 2	
	N_tel	Level3	
«Fournisseur »	N_four	Level 1	Hierarchy 1 : H1 = N1 à N2 .
	Nom_prenom		
	Ville	Level 2	
«Commande»	Id_com	Level 1	Hierarchy 1 : H1 = N1 à N2 .
	Etat	Level 2	

Table 7.1 – List of levels and hierarchies of each dimension.

7.3 The listing of the cubes

In the board below, we will train the listing of the cubes at achieve, For each cube we will give the dimensions participating so than the measurable present in these cubes.

Cube	The measurable	the dimensions
Vente	-Benefice -Prix_a_payer	Produit
		Facture
		Depot
		Point_vente
		Temps
Approvisionnement	-Montant_a_pay -Montant_verse -Montant_reste	Produit
		Commande
		Temps
		Depot
Stock	-Qte_totalST -Nb_fact -Ttc_ST -prix_achat -Htt_St -Qte	Depot
		Temps
		Type_stock

Table 7.2 – List of multidimensional cubes.

7.4 Presentation of the cubes dimensional

To of reply all the needs of the company we have designed three cubes **OLAP**. We will represent each of the cubes with the table of made at which he belongs, detected measures as well as dimensions.

Cube "Analysis Sale" :

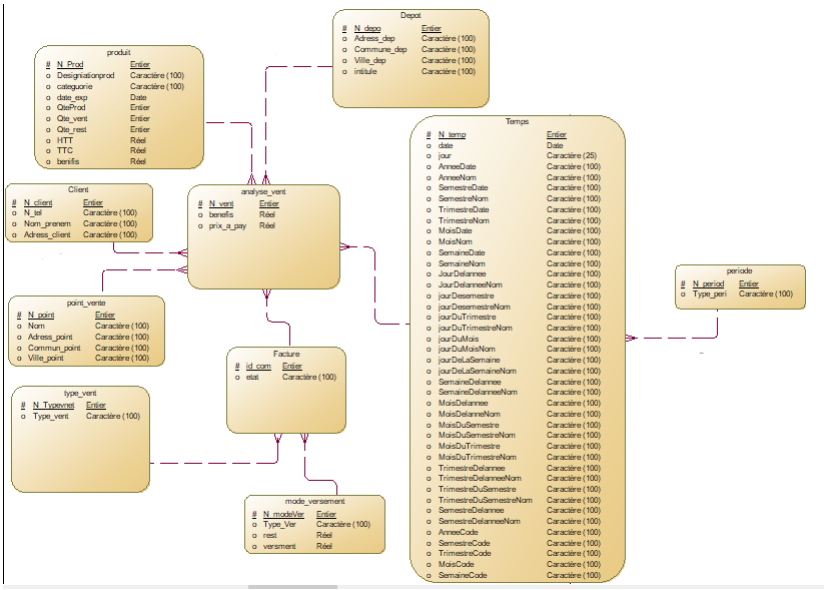


FIGURE 7.1 – Dimensional Cube "Sale".

Cube "Supply Analysis" :

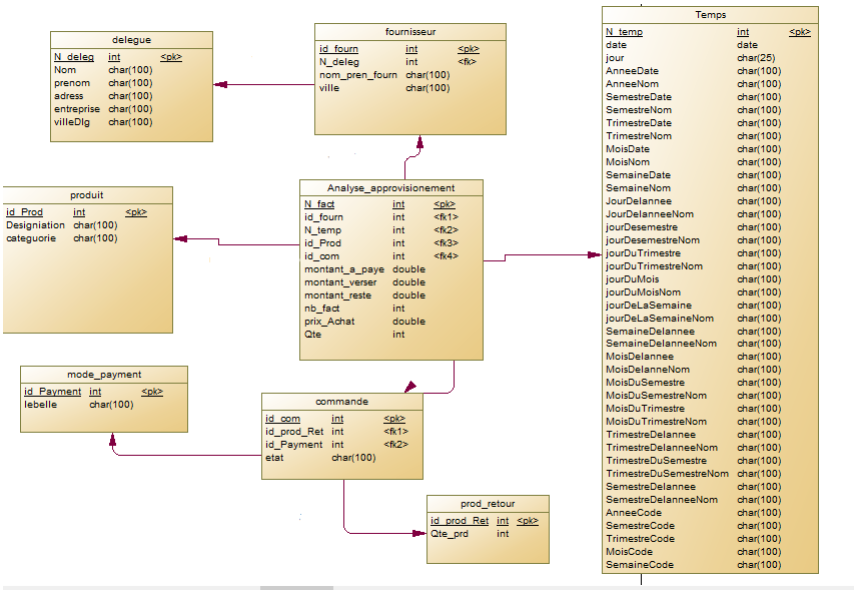


FIGURE 7.2 – Dimensional Cube "Purchase".

Cube "Stock analysis" :

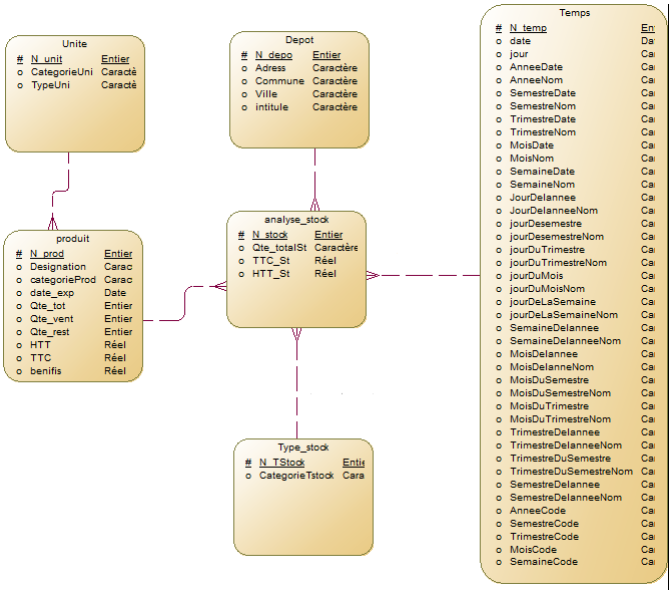


FIGURE 7.3 – Dimensional Cube "Stock".

observation :

Concerning the dashboard by not using the visualization tools so by programming it, it goes the same or even In the implementation chapter.

7.5 Conclusion

The design of the cubes dimensional must to permit the users an exploitation of data of way simple and intuitive, an analysis interactive of the data. These cubes are going to permit the makers having the state of the activity in time real.

In this chapter we have defined the levels and hierarchies of each dimension, and the different dimensional cubes by specifying the participating dimensions.

Chapitre

8 implementation

Sommaire

8.1 The architecture of the solution . . . 73

8.2 Resources used . . . 75

8.2.1 Material resources . . . 75

8.2.2 Software resources . . . 76

8.2.3 Comparative study of Open Source BI tools . . . 76

8.3 Implementation of the solution . . . 77

8.3.1 Implementation of the ETL . . . 77

8.3.2 Implementation of the data analysis and reporting component . . . 79

8.3.3 Choice of method . . . 82

8.4 The multiple linear model . . . 82

8.4.1 Writing the model . . . 82

8.4.2 Principle of the calculation . . . 82

8.5 Conclusion . . . 85

After having identified the users' needs, classified them and designed the solution composed of four steps **(Datawarehouse, ETL, OLAP cubes and dashboard)**, in this chapter we discuss the implementation of the system, starting with the tools used, the architecture adopted and a brief presentation of the solution implemented.

8.1 The architecture of the solution

The architecture of the solution is shown in the following **figure8.1** :

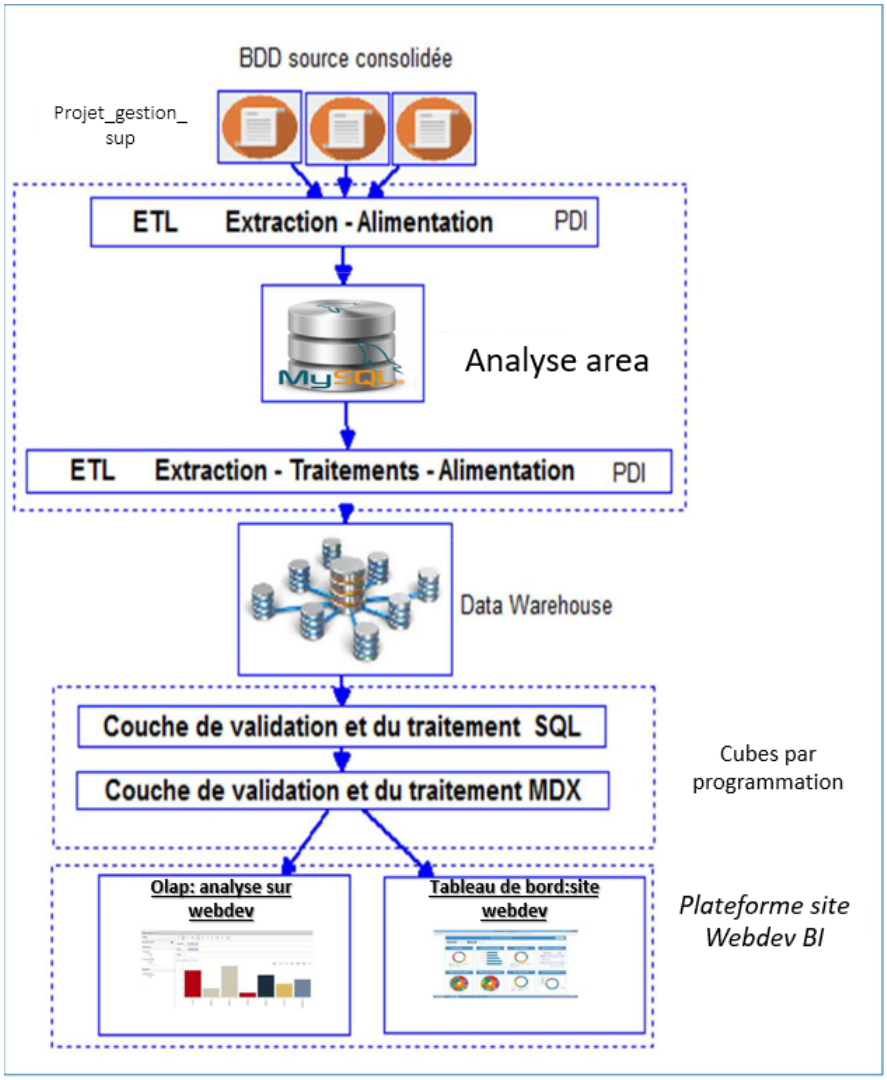


FIGURE 8.1 – Overall solution architecture.

The architecture of the restitution part is represented in the following **figure8.2** :

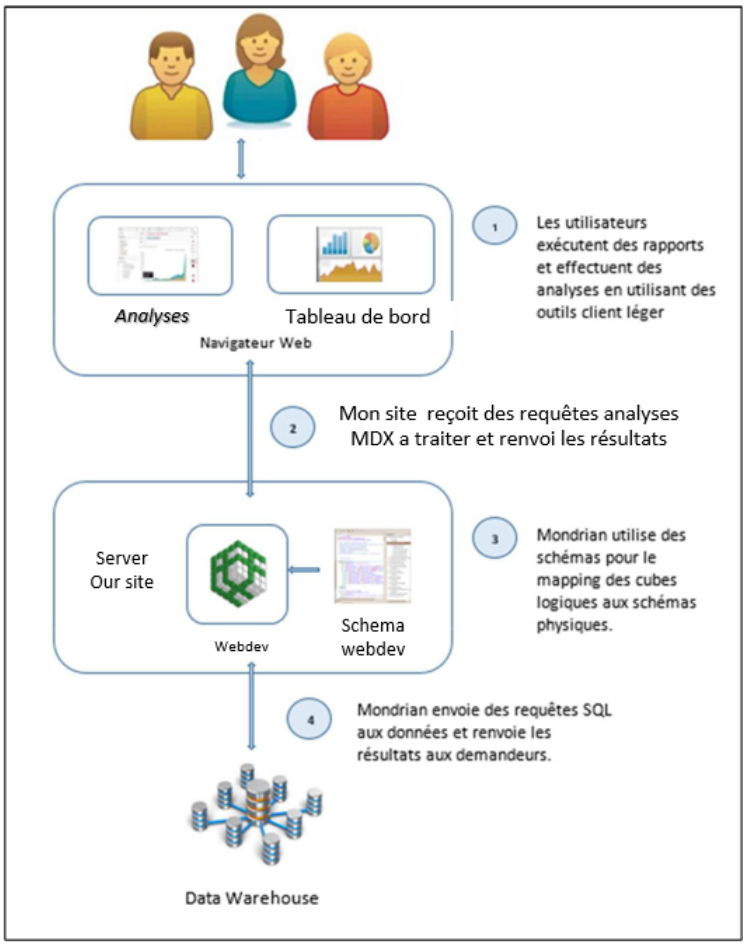


FIGURE 8.2 – The architecture of the restitution part.

8.2 Resources used

8.2.1 Material resources

For the realization of the project, we used our own resources whose characteristics are as follows :

— Windows 10, i5, 4GB of RAM.

8.2.2 Software resources

Tool	Version	Description
MySql	9.6	DBMS used for data storage.
Talend	ce 6.1	A feeding tool transformation
Talend	cd 6.1	A plugin that can be integrated into the BI server, which allows you to visualize and analyze the Talend data from an OLAP cube in a table ra drag and drop graph from the defined OLAP model. this uses the Mondrian engine and the pilot OLAP4J to retrieve the information at from a data source and the restore in the browser or export them.
webdev	17	plugin,which adds new types of graphic illustrations at web Analytics.
webdev	17	allows you to deliver dashboardswebdev of a professional quality, minimizing the complexity associated with code writing.
webdev par Programmation	17	a cube-design application Mondrian.The application is overlay compared to manual editing XML files : each can be modified using a form basic.
webdev	17	to create composite pages in Web mode.

Table 8.1 – Tools used during the development of the solution.

8.2.3 Comparative study of Open Source BI tools

After a comparative study between the three best-known open source tools in the world of business intelligence and reporting, we were able to synthesize that the three tools provide a rich set of ready-to-use functionalities for the company.

Significant differences can be found in the presentation of reports, or in the availability of a report server. webdev stands out as more than just a reporting tool, with a complete suite of components (data mining and integration).

Here is a comparative table of the most popular open source BI tools for example **figure8.3** :

	PowerAMC	Talend	Webdev mon site web
ETL	✓	✓	×
REPORTING	×	×	✓
Tableau de bord	×	×	✓
Métadonnées et requêtage ad_hoc	✓	×	✓
Analyse Multidimensionnelle (olap)	×	×	✓
Serveur BI	×	×	✓

Figure 8.3 – Comparison table of the best open source BI tools.

The company did not require any tools for the implementation of the solution, so we had to choose a feature-rich tool at a lower cost. And since the webdev suite is a complete Open Source suite, we have chosen it for the implementation of the BI system.

8.3 Implementation of the solution

8.3.1 Implementation of the ETL

For the implementation of the **ETL**, we used the Talend tool which allows, thanks to a graphical tool based on steps, to :

- Connect to databases.
- Create processes consisting of data imports and exports without programming.
- Design and execute handling and processing operations on these data.

We proceeded as follows :

8.3.1.1 Definition of connections to data sources

All transformations included in the **ETL** define two MySql database connections : One is called **project_management_sup database**, which connects to the analysis database that will be populated with data extracted from flat files. The other is called **DWH** which connects to the Data Warehouse (the target).**figure8.4**

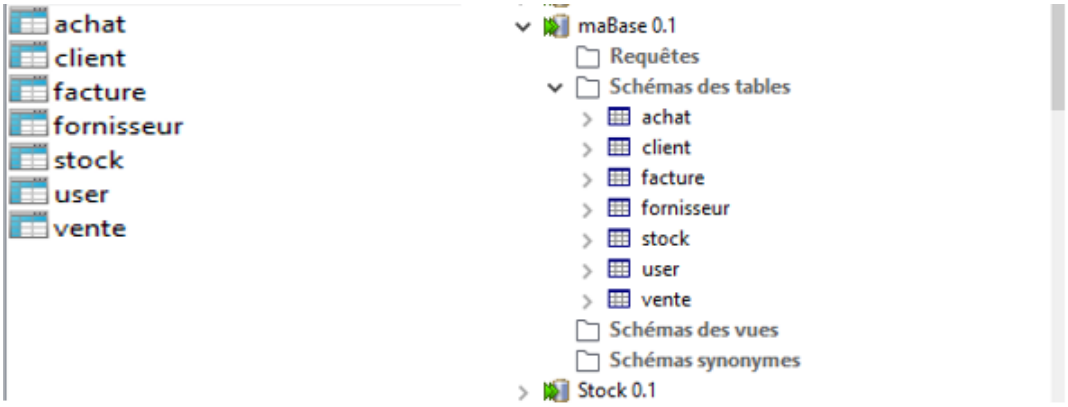


FIGURE 8.4 – Connection to data sources.

The actual connection to the database is only made at runtime.

8.3.1.2 Loading the "analysis" database allows you to

- Store data extracted from source databases
- Do operations on these data[**Kimball.2010**]

The data in the "analysis" area is destroyed before the next Data Warehouse loading is started, so it has a longer or shorter life span. [Steiner.2013]

The data source for the data in the "analysis" Area is UNL files.like the figure **figure8.5**

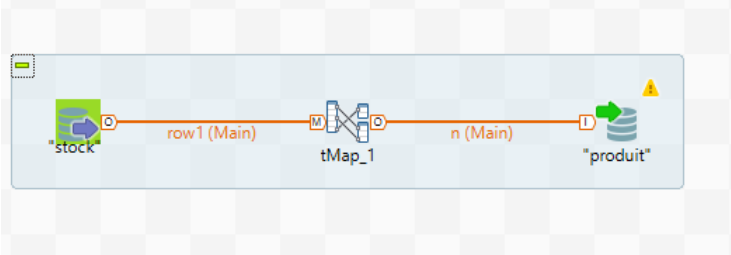


FIGURE 8.5 – Loading analysis Area.

The same goes for the other tables.

8.3.1.3 Cleaning of the "analysis" data Area

The reason for adding steps to clean up data and comply with business requirements before loading the data warehouse is the simple fact that there is no organization in the world that does not have a data quality problem.

8.3.1.4 Loading the DW

We used sequential tasks to synchronize transformations and send success or failure notifications by email.

The loading of the dimensions starts first. Once completed, the loading of the de facto tables begins. as the **figure8.6**

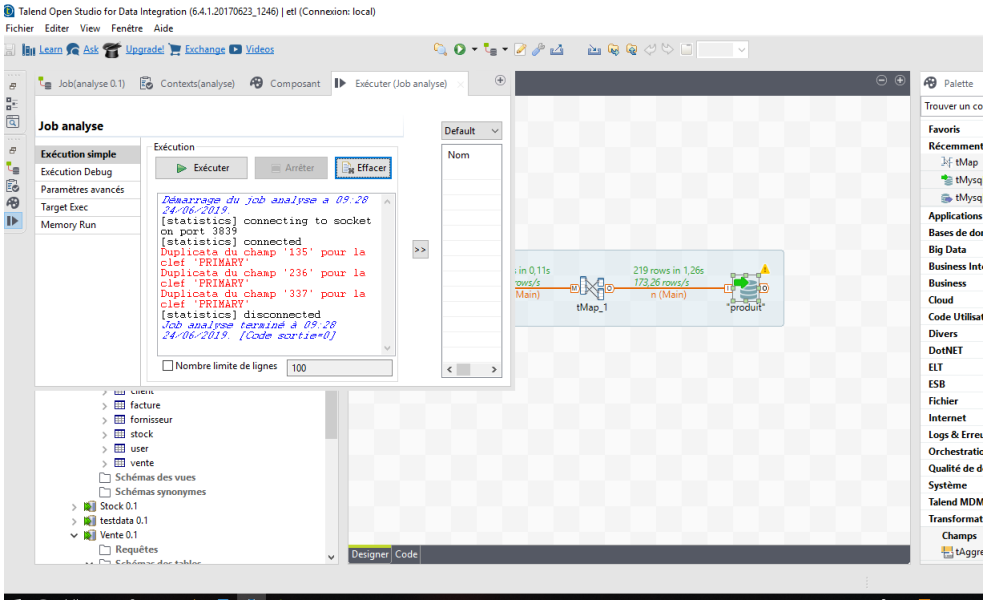


FIGURE 8.6 – ETL.

8.3.2 Implementation of the data analysis and reporting component

8.3.2.1 Creating the OLAP schema

In this section, we have created an **OLAP** schema for the data warehouse using the PowerAMC tool that requires a connection to the DWH (MySql DBD).

Our Webdev site uses the **ROLAP** (Relational Online Analytical Processing) engine, called by **Mondrian** to convert MDX (abbreviated to MultiDimensional eXpressions) queries to SQL queries.

We have created cubes that represent fact tables and include dimensions in the schema with hierarchies to increase the performance of multidimensional navigation.

The schema must be saved and published on the server of our BI Server website. [9]

8.3.2.2 Data analysis and restitution component

After the publication of the site on the BI server of the web, We always used for the analysis of the data the **Webdev** tool on our site, enriched with a plugin.

The tool allows you to view and analyze the data of an OLAP cube in a pivot table or drag and drop graph from the defined OLAP model.

On our site offers almost all the functions available :

- Data positioning in the form of a "Pivot table" with possible axis permutation.
- Data filtering via a selection box.

9. Mondrian is an OLAP (Online Analytical Processing) engine written in Java by Julian Hyde that allows the design, publication and query of multidimensional cubes. The Mondrian community is hosted and supported by the publisher Pentaho.

- Automatic or manual execution of the MDX request (useful when you want to configure all axes before you even retrieve the data).
- Showing/hiding the side panel and column selection area + rows + filters.
- Showing/hiding empty lines.
- Displaying the MDX request.
- Graphics generation engine, in which you can enable/disable the display of certain series : ideal for making comparisons.

Example of purchase analyses .like the **figure8.7**.

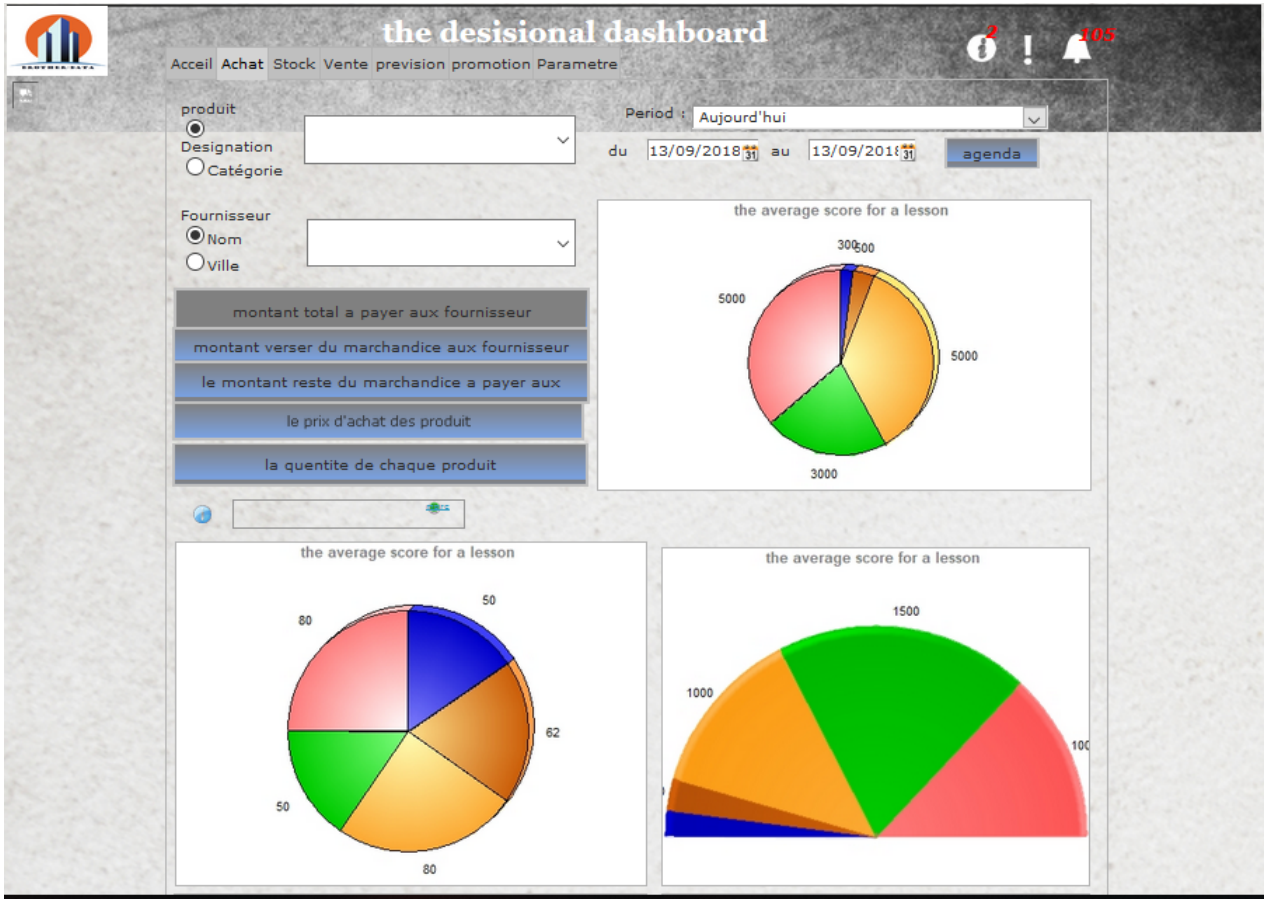


FIGURE 8.7 – Example of purchase analyses (total amount to be paid to suppliers by product).

Example of a graph as the **figure8.8**.

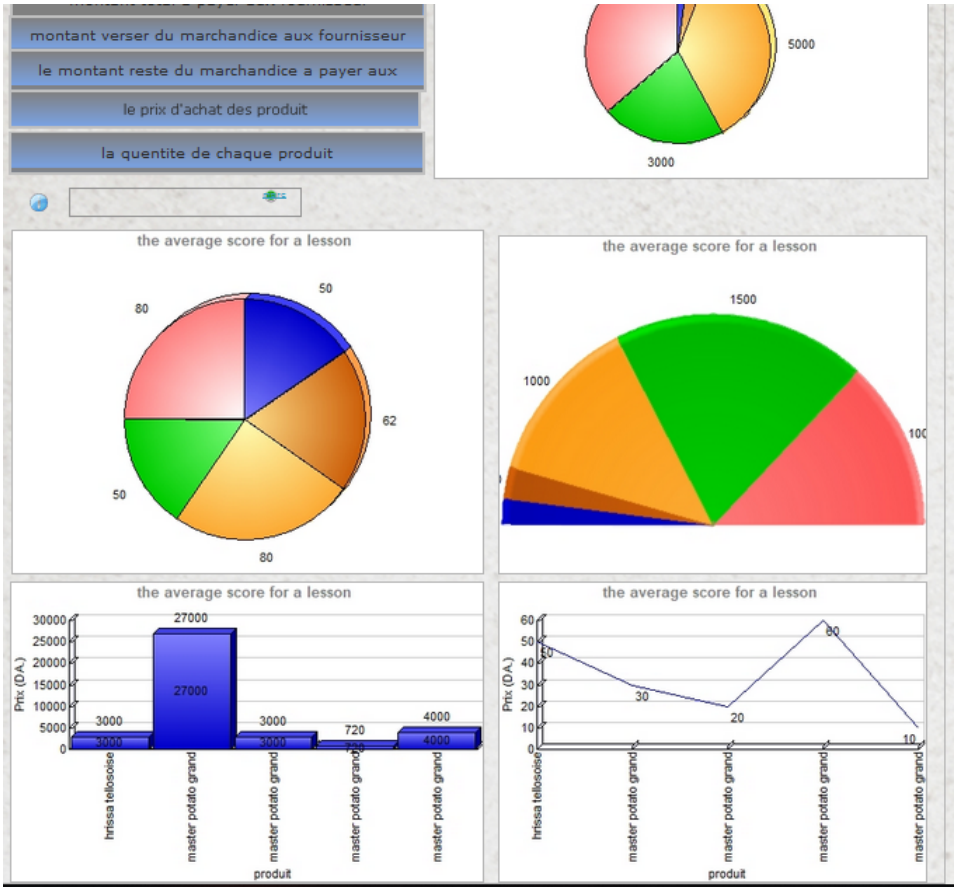


FIGURE 8.8 – Example of a graph (the amount of the remaining merchandise to be paid to suppliers).

8.3.2.3 Dashboard

using multiple linear regression in the interface or flight prediction as shown in the **figure 8.7**.

8.3.2.4 Regression equation

the regression equation is written :

$$Y_i = a_0 + a_1 * x_1 + ... + a_p * x_{ip} + \epsilon_i \qquad , i = 1...p \tag{8.1}$$

i is the error of the model, it expresses, or summarizes, the missing information in the linear explanation of the values of y from x j a0, a1, ..., ap are the coefficients (parameters) of the model to be estimated

8.3.2.5 The modelling processes

— Estimate the values "a" by using the data

— Evaluate the accuracy of these estimators.
— Measures the explanatory power of the model .
— Evaluate the influence of variables in the model .
— Select the most relevant variables.
— Evaluate the quality of the model when predicting .
— Detect observations that may overly influence the results .

Advantages and disadvantages of the technique

Advantages

— The applications of linear regression are extremely numerous.
— Explanatory variables are both quantitative and qualitative variables.

Disadvantages

— It does not address highly non-linear problems.
— Too much calculation : in the case where the observation base is very large.

8.3.3 Choice of method

We have chosen the multiple linear regression method because it seems to us to be the most appropriate for this type of problem.

8.4 The multiple linear model

8.4.1 Writing the model

Let us consider the variables Y and (X1,…..,Xk) observed on a sample size n. For observation number i we have(Yi.Yi1……Yik). We note by X the matrix (n,k) of values of the variables(X1,……Xk). Each line of X is made up of the values Xk taken during observant n>k).

$$X = \begin{bmatrix} X11 & .. & X1k \\ X11 & .. & X1k \\ . & .. & . \\ . & .. & . \\ Xn1 & .. & Xnk \end{bmatrix}$$

8.4.2 Principle of the calculation

The multiple model is written :

$$\begin{cases} Y_i = \sum_{h=1}^{k} b_h X_{ih} + u_i & i \in [1;n] \\ E(u_i) = 0 & V(u_i) = \sigma^2 \\ E(u_i, u_l) = 0 & (i \neq l) \end{cases}$$

If we note :

$$Y = \begin{bmatrix} Y_i \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} \quad u = \begin{bmatrix} u_i \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{bmatrix} \quad b = \begin{bmatrix} b_i \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{bmatrix}$$

The model is written in matrix form :

$$\begin{cases} Y = Xb + u \\ E(u) = 0 \\ V(u) = I_n \ \sigma^2 \end{cases}$$

Least Squares Estimation

We are looking for an estimate $\widehat{b}$ of b who minimizes ssr(b) :

$$\text{ssr(b)=ssr(b}_i, \ldots, b_k) = \sum_{i=1}^{n} u_i^2$$

$$\min_b \ \sum_{i=1}^{n} u_i^2 = \min_b \sum_{i=1}^{n} (Y_i - \sum_{l=1}^{k} b_l X_{il})^2$$

We can write ssr(b)

:

$$ssr(b) = (Y - bX)^t (Y - bX) = Y^t Y - 2b^t X^t Y^t + b^t X^t Y b$$

The application of the least squares criterion, consisting in minimizing the latter, leads to normal equations in matrix form :

$$\text{-}2^t XY \ + \ X^t X\widehat{b} = 0$$

Hence the least squares estimator of :

$$\widehat{b} = (X^t X)^{-1} X^t Y$$

For the creation of the dashboards, we used the **webdev** tool and by programming to perform the visualization tasks.like the**figure8.9** , the**figure8.10** and the**figure8.11**.

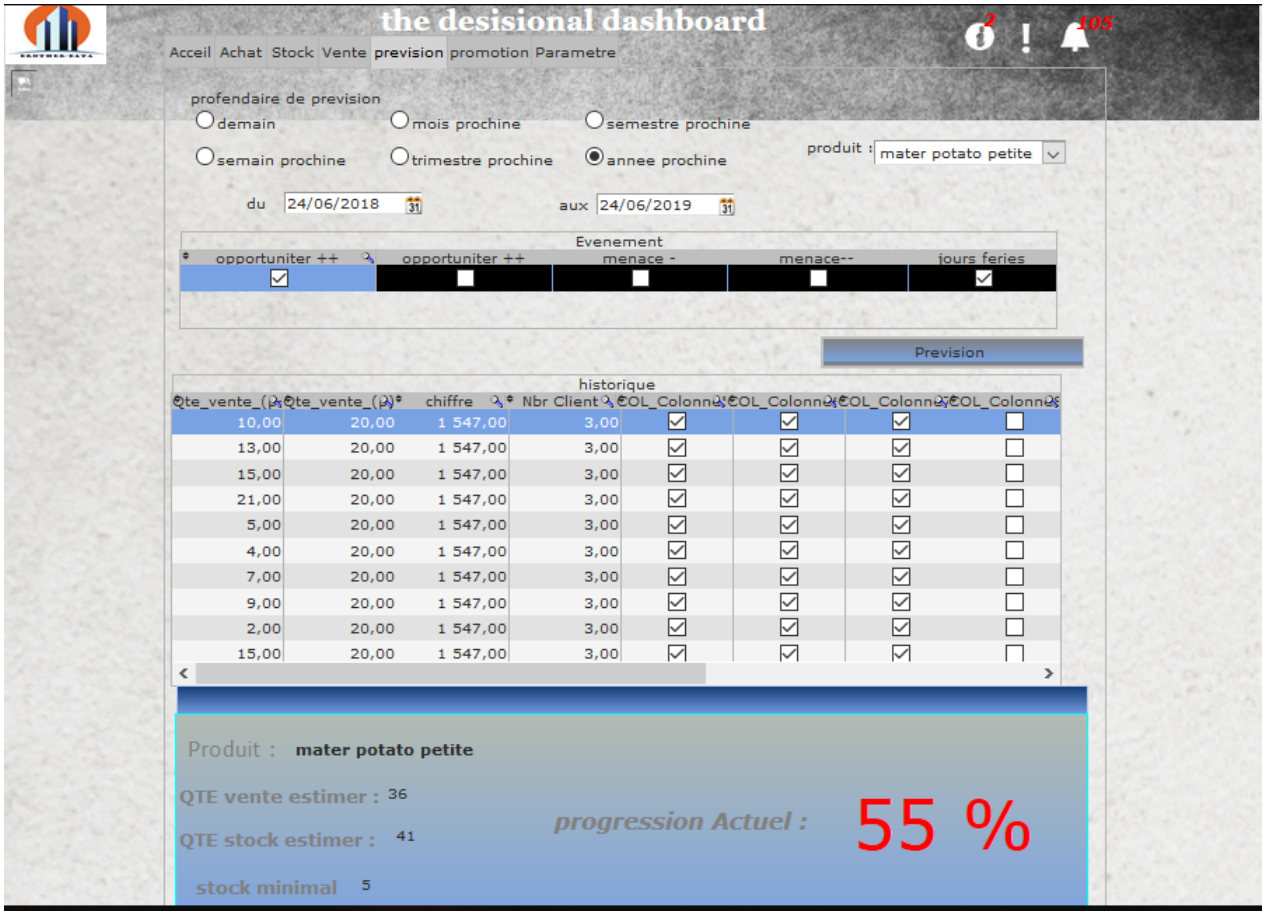


FIGURE 8.9 – Dashboard of the physical achievements of the forecast .

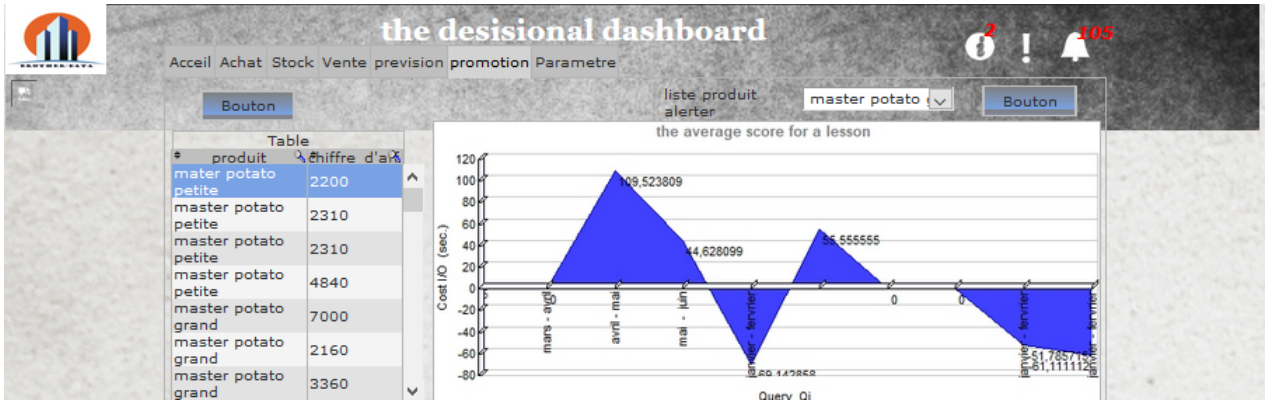


FIGURE 8.10 – Dashboard of the physical achievements of the promotion(tendence).

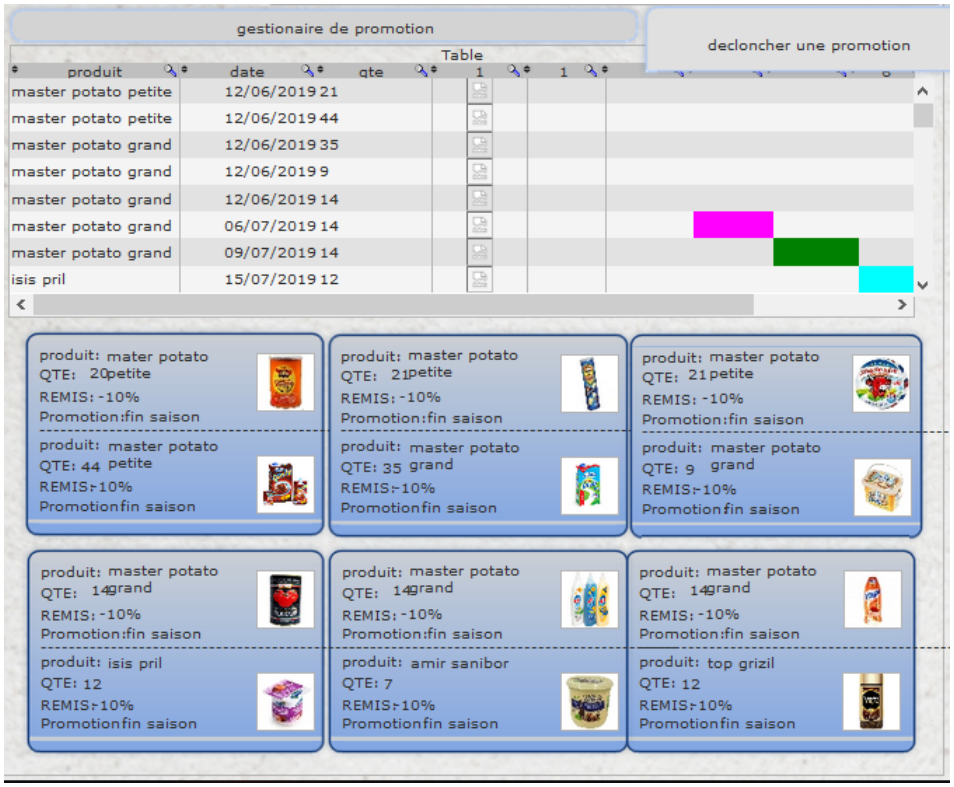


FIGURE 8.11 – Dashboard of the physical achievements of the promotion.

Security of the solution

The default web interface of on webdev, gives very simple access to all available actions and administration functions. Security is widely implemented at all expected levels : access to documents, rights of action on documents and also data security according to the role of the end user. The notion of the role is present in this platform and in the OLAP schema construction tool.

8.5 Conclusion

Through this chapter, we have explained the different points taken into account for the implementation of the decision support system as well as the results obtained at the level of this phases.

Our solution was developed with **Talend** and **PowerAMC** ,**Navicat** ,and our own website on **Webdev**, each software being designed for a particular task of our system.

Chapitre

9

Overall conclusion

Sommaire

9.1 Overall conclusion 87

9.1 Overall conclusion

In recent years, business intelligence systems have been high on the corporate priority list. In order to be attentive to their performance, and always in line with their customers' expectations. This requires companies to have a global view of their business, identify their weaknesses and detect opportunities, by collecting and analyzing the information stored. A requirement in a rapidly changing global market. To provide company managers with access to reliable and relevant information to make strategic management decisions more intelligently and give them the ability to analyze the company's activity. Who initiated the project to set up a decision support system that was the subject of our final year project. We started the project by doing a bibliographic search, which was a very important phase. It allowed us to get an idea of the decision-making systems and to familiarize ourselves with the elements necessary for their implementation. At the end of this In this phase, we opted for the Kimball approach, which considers the Data Warehouse as a set of DataMarts that are consistent with each other. Its advantage is the speed of implementation of a business intelligence solution adapted to the needs described. We then carried out a complete and detailed study on the state and procedures of decision-making within the company as well as an identification of users and identify their needs. Through interviews, supplemented by questionnaires and observations. Based on this study and following Kimball's approach, we have identified three topic-oriented DataMarts. One Datamart for the sales activity, one Datamart for the purchasing activity and another for the stock activity.

The design phase began with the design of the storage area. In this section, we were able to identify all the dimensions, facts and measurements of the data warehouse as well as the multidimensional (star-shaped) schemas of each activity. Then we moved on to the design of the feed zone which consisted in extracting the data from the MySQL transactional database, using the WampServer, PowerAMC tools). These services include a rich set of integrated tasks and transformations and graphical tools that allow you to extract and

Transform data from a possible source without writing a single line of code (use only a loading script). The loading of the cubes and dimensions of our DataMarts, has been done with MySQL Server and Talend which allows you to design, create and manage multidimensional structures. At this stage, the

cubes are loaded and ready to be interrogated. We can now deploy our system, providing users with a tool that offers a wide range of data analysis and presentation options (our webdev17 website). By setting up this platform, we were able to analyze and collect all the information necessary for decision-making and allow users to have reports, supplies and sales promotions of their activity in real time.

At the end of this project, we were able to achieve the objectives we had previously defined.

Part IV

Abriviation and Bebliography

Abreviation :

ERP : Enterprise Resource Planning

BDD : Data Bases

DW : Data Warehouses

DMS : Database Management System

DM : Data Marts

APL : Programming Language

BI : Business Intelligence

ETL : Extracts,Transformed and Loaded

OLAP : On Line Analytical Processing

MOLAP : Multidimensional OlAP

ROLAP : Relational OLAP

HOLAP : OLAP Hybrid

OLTP : Online Transaction Processing

RDBMS : Relational Database Management System

DBMS : Database Management System

DSS : Decision Support System

SID : Decision-making Information System

FCA : Formal Concept Analysis

CBR : Case based reasoning

MDX : Multidimensional Expressions

Bibliographie :

[Burquier, 2009] : Bertrand Burquier, « *Business Intelligence avec SQL Server 2008- Mise en œuvre d'un projet décisionnel* », Editions DUNOD ,2009.

[Kimball, 1996] : R. Kimball ; « *Entrepôts de données : Guide pratique du concepteur de Data Warehouse* », Wiley Computer Publishing 1996.

[Kimball, al, 2007] : R. Kimball, M. Ross et W. Thornthwaite, « *Le Data Warehouse – Guide de conduite de projet* », Editions Eyrolles, 2007-06-11.

[OSMANI, 2009] : NAILI D., OSMANI F., Conception et réalisation d'un portail décisionnel pour L'Entreprise ELSECOM, Ecole Nationale Supérieure d'Informatique, Alger, 2009.

[Yahi, 2012] : MERAGHNI O., Yahi I., Conception et réalisation d'un DATA WAREHOUSE pour la mise en place d'un système décisionnel au sein du groupe TOYOTA ALGERIE, Ecole Nationale Supérieure d'Informatique, Alger, 2012.

[Bill Inmon (1996)] **:** [7]

[Inmon, 2005] : [8]

[Marcel, 98] : [9]

[GRIM, 2009] : Architecture olap [10]

Webographie

[1] Nodesway, http://www.nodesway.com/tableaudebord/indicateur-performance.htm

[2] Fernandez Alain, Piloter la performance, http://www.piloter.org/business-intelligence/reporting.htm

[3] Fernandez Alain, Piloter la performance, http://www.piloter.org/business-intelligence/tableau-de-bord-business-intelligence.htm

[4] Fernandez Alain, Piloter la performance, http://www.piloter.org/business-intelligence/business-intelligence.htm

[5] Developpez.com http://blog.developpez.com/jmalkovich/p8718/modelisation/modele_en_etoile_ou_en_flocons

[6] Developpez.com http://grim.developpez.com/cours/businessintelligence/concepts/conception-datawarehouse/#LV-A

[7] datawarehouse.ppt *www.cse.iitb.ac.in/dbms/Data/Talks/krithi-talk-impact.ppt*

[8] structure data Warehouse.pdf https://www.ismll.uni-hildesheim.de/lehre/.../Inmon-vs-Kimball.pdf

[9] *www.lsis.org/espinasseb/Supports/DWDM/3-OLAP-4p.pdf*

[10] *igm.univ-mlv.fr/~dr/XPOSE2009/informatique_decisionnelle_olap/colap.html*