**PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA**

**MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH**

**MOHAMED BOUDIAF UNIVERSITY - M'SILA**

**FACULTY: Mathematics and Informatics**
**Department of computer science**
**N°: ……**

**DOMAIN: Mathematics and Informatics**
**FIELD: Computer Science**
**SUB-FIELD: Software Engineering**


A Dissertation in Fulfillment

**For the Requirements of the Degree of Master**

**By**: MAHDI Chawki

**SUBJECT**

# Data Warehouse CRM Oriented

**Defended publicly on:12/06/2017**

**Board of Examiners:**

| | | |
|---|---|---|
| **……..…** | **M'silaUniversity** | **Chairman** |
| **BOUNIF Mohamed** | **M'silaUniversity** | **Supervisor** |
| **……..…** | **M'silaUniversity** | **Examiner** |
| **……..…** | **M'silaUniversity** | **Examiner** |

**Promotion: 2016/2017**

## Acknowledgemnts

I would be like to thank Dr. Bounif Mohamed my supervisor for his support and guidance through this work. I would like to thank all the professors of Advanced Information Systems department for their valuable advices and encouragement. Special thanks go to Dr. Ouness Bilal for his help and support in solving many problems associated with this work.

Finally, I am very grateful to my family for their unlimited love and support during my studies. And wed don’t forget to give Nessplus Telecom special thanks for their help and their trust.  This work is dedicated to them and all my friends who have supported me.

# Table des matières

## List of Figures



## List of Tables:

# Introduction

In a 1958 article, IBM researcher Hans Peter Luhn used the term business intelligence, he defined intelligence as:" the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal ".

For companies maintaining direct contact with large numbers of customers, however, a growing number channel-oriented applications (e.g. e-commerce support, call center support …etc.) create a new data management challenge: that is effective way of integrating enterprise applications in real time. And to learn from the past and forecast the future, many companies are adopting Business Intelligence (BI) tools and systems. Companies have understood the importance of enforcing achievements of the goals defined by their business strategies through business intelligence concepts.

Since the beginning of BI term, there are a several different definitions. Although, a lot of different terms refers to BI which can be defined as set of strategies and technologies and processes provides technics and methods used by enterprises to support the collection, data analysis … etc. to help and generate a useful knowledge for better business decision-making.

In this changing world and with the incredible growing the amount of data today, each enterprise needs to know about the situation of their customers and what they need and what's their interests, what it should to do for satisfy them and increase their profits from customers. For this the CRM approach (**C**ustomers **R**elationships **M**anagement) come to resolve this problem, CRM consists in gathering data and information for design a business strategy to reduce cost and increase economic gains by strengthen customer satisfaction & loyalty and improve customers relationship with enterprises.

Many enterprises nowadays store all of the daily transaction with their customers, but they don't have awareness about the great importance of these data, and how these data can help them to improve and increase their economic gains, Although, they don't have the suitable way to exploit it.

Even if they knew the importance of data which they collected, they would stay unaware about how to collect data and how to put it and the most difficult one is how to extract knowledge from these data

Let’s take an example about that, Nessplus Telecom is a telecommunication company and they have a lot of customers in different regions, which makes it very difficult to collect data from the sale points & extract knowledge from this data, and of course how to evaluate their work and to study strategies if is it efficiency or not, this cost a lot of time and efforts to collect and cleaning data in one form, especially that they come from heterogenous sources.

This is what data warehouse found to solve it, which is a repository and container of data, design to store and collect and extract large amount of data from heterogenous sources (csv, excel, flat files …etc.) and creating a historical record of all customer interactions, the abilities to view and manipulate set data warehouses apart from other computer systems. Constantly extracting knowledge about customers reduces the need for traditional marketing research tools such as customer surveys and focus groups

The CRM purposes to authorize the manage customer needs & anticipations more effectively in order to increase gains profits by extract knowledge from customer gathered information across many CRM applications for achieve that purpose.

In this thesis we have five chapters, in the first chapter we are going to introduce business intelligence technologies and what is the main goal from this approach and how it can help, and explain the components and the architecture of BI.

In the second chapter, we will talk about data warehouse and how is solved a big problem, and described process of the creation of data warehouse, how to extract data from heterogenous systems and sources and transform it into one homogeny repository.

In the third chapter, we will introduce two methods for extracting knowledge from data warehouse to help decision-makers, which are data mining and business reporting, and what are the most software’s and algorithms use it for that purpose.

In the fourth chapter, we will talk about the main idea of CRM’s thesis, what is CRM and how it can help decision-makers to know their customers, and how to satisfy them, and of course how to increase gains.

In the last chapter of implementation, we will take a real example about Nessplus company, and we are going to apply CRM approach from a big unrelated data to create data warehouse and extracting knowledge and help them to increase their profits

# CHAPTER 1

# Business Intelligence

In the first chapter, we talk about the Business intelligence (BI) has two basic different meanings related to the use of the term intelligence. The primary, less frequently, is the human intelligence capacity applied in business affairs/activities. Intelligence of Business is a new field of the investigation of the application of human cognitive faculties and artificial intelligence technologies to the management and decision support in different business problems.

The second relates to the intelligence as information valued for its currency and relevance. It is expert information, knowledge and technologies efficient in the management of organizational and individual business. Therefore, in this sense, business intelligence is a broad category of applications and technologies for gathering, providing access to, and analyzing data for the purpose of helping enterprise users make better business decisions. The term implies having a comprehensive knowledge of all of the factors that affect the business. It is imperative that firms have an in depth knowledge about factors such as the customers, competitors, business partners, economic environment, and internal operations to make effective and good quality business decisions. Business intelligence enables firms to make these kinds of decisions.

A specialized field of business intelligence known as competitive intelligence focuses solely on the external competitive environment. Information is gathered on the actions of competitors and decisions are made based on this information. Little if any attention is paid to gathering internal information.

## 1. What is Business Intelligence?

There are many definitions for Business Intelligence, or BI, but maybe the closest one or the clear one is: **Business Intelligence** (**BI**) are the set of strategies, processes, applications, data, technologies and technical architectures which are used to support the collection, analysis, presentation and dissemination of business information [1]. to make it easier **BI** is about gathering and delivering relevant and reliable information to the decision-makers at the right time to help them achieving their goals by making the best decision in short time.

## 2. Why Business Intelligence? [2]

Business Intelligence enables organizations to make well informed business decisions and thus can be the source of competitive advantages. This is especially true when firms are able to extrapolate information from indicators in the external environment and make accurate forecasts about future trends or economic conditions. Once business intelligence is gathered effectively and used proactively then the firms can make decisions that benefit the firms. The ultimate objective of business intelligence is to improve the timeliness and quality of information. Timely and good quality information is like having a crystal ball that can give an indication of what's the best course to take. Business intelligence reveals:

- The position of the firm as in comparison to its competitors
- Changes in customer behavior and spending patterns
- The capabilities of the firm
- Market conditions, future trends, demographic and economic information
- The social, regulatory, and political environment
- What the other firms in the market are doing [2]

## 3. Components of BI [2]

### 3.1. OLAP (On-line analytical processing):

It refers to the way in which business users can slice and dice their way through data using sophisticated tools that allow for the navigation of dimensions such as time or hierarchies. Online Analytical Processing or OLAP provides multidimensional, summarized views of business data and is used for reporting, analysis, modeling and planning for optimizing the business. OLAP techniques and tools can be used to work with data warehouses or data marts

designed for sophisticated enterprise intelligence systems. These systems process queries required to discover trends and analyze critical factors. Reporting software generates aggregated views of data to keep the management informed about the state of their business. Other BI tools are used to store and analyze data, such as data mining and data warehouses; decision support systems and forecasting; document warehouses and document management; knowledge management; mapping, information visualization, and dash boarding; management information systems, geographic information systems; Trend Analysis; Software as a Service (SaaS).

### 3.2.Advanced Analytics:

it is referred to as data mining, forecasting or predictive analytics, this takes advantage of statistical analysis techniques to predict or provide certainty measures on facts.

Corporate Performance Management (Portals, Scorecards, Dashboards): this general category usually provides a container for several pieces to plug into so that the aggregate tells a story. For example, a balanced scorecard that displays portlets for financial metrics combined with say organizational learning and growth metrics.

### 3.3. Real time BI:

It allows for the real time distribution of metrics through email, messaging systems and/or interactive displays.

### 3.4.Data Warehouse and data marts:

The data warehouse is the significant component of business intelligence. It is subject oriented, integrated. The data warehouse supports the physical propagation of data by handling the numerous enterprise records for integration, cleansing, aggregation and query tasks. It can also contain the operational data which can be defined as an updateable set of integrated data used for enterprise wide tactical decision-making of a particular subject area. It contains live data, not snapshots, and retains minimal history. Data sources can be operational databases, historical data, external data for example, from market research companies or from the Internet), or information from the already existing data warehouse environment. The data sources can be relational databases or any other data structure that supports the line of business applications. They also can reside on many different platforms and can contain structured information, such as tables or spreadsheets, or unstructured information, such as plaintext files or pictures and

other multimedia information. A data mart as described by (Inmon, 1999) is a collection of subject areas organized for decision support based on the needs of a given department. Finance has their data mart, marketing has theirs, and sales have theirs and so on. And the data mart for marketing only faintly resembles anyone else's data mart. Perhaps most importantly, (Inmon, 1999) the individual departments own the hardware, software, data and programs that constitute the data mart. Each department has its own interpretation of what a data mart should look like and each department's data mart is peculiar to and specific to its own needs. Similar to data warehouses, data marts contain operational data that helps business experts to strategize based on analyses of past trends and experiences. The key difference is that the creation of a data mart is predicated on a specific, predefined need for a certain grouping and configuration of select data. There can be multiple data marts inside an enterprise. A data mart can support a particular business function, business process or business unit.

A data mart as described by (Inmon, 1999) is a collection of subject areas organized for decision support based on the needs of a given department. Finance has their data mart, marketing has theirs, and sales have theirs and so on. And the data mart for marketing only faintly resembles anyone else's data mart.

BI tools are widely accepted as a new middleware between transactional applications and decision support applications, thereby decoupling systems tailored to an efficient handling of business transactions from systems tailored to an efficient support of business decisions. The capabilities of BI include decision support, online analytical processing, statistical analysis, forecasting, and data mining. The following are the major components that constitute BI.

### 3.5. Data sources:

can be operational databases, historical data, external data for example, from market research companies or from the Internet), or information from the already existing data warehouse environment. The data sources can be relational databases or any other data structure that supports the line of business applications. They also can reside on many different platforms and can contain structured information, such as tables or spreadsheets, or unstructured information, such as plaintext files or pictures and other multimedia information. [3]


*Figure 1 The main components of a BI*

## 4. Benefits of BI

BI provides many benefits to companies utilizing it. It can eliminate a lot of the guesswork within an organization, enhance communication among departments while coordinating activities, and enable companies to respond quickly to changes in financial conditions, customer preferences, and supply chain operations. BI improves the overall performance of the company using it.

Information is often regarded as the second most important resource a company has (a company's most valuable assets are its people). So when a company can make decisions based on timely and accurate information, the company can improve its performance. BI also expedites decision-making, as acting quickly and correctly on information before competing businesses do can often result in competitively superior performance. It can also improve customer experience, allowing for the timely and appropriate response to customer problems and priorities.

The firms have recognized the importance of business intelligence for the masses has arrived. Some of them are listed below.

- With BI superior tools, now employees can also easily convert their business knowledge via the analytical intelligence to solve many business issues, like increase response rates from

direct mail, telephone, e-mail, and Internet delivered marketing campaigns.

• With BI, firms can identify their most profitable customers and the underlying reasons for those customers' loyalty, as well as identify future customers with comparable if not greater potential.

• Analyze click-stream data to improve ecommerce strategies.

• Quickly detect warranty-reported problems to minimize the impact of product design deficiencies.

• Discover money-laundering criminal activities.

• Analyze potential growth customer profitability and reduce risk exposure through more accurate financial credit scoring of their customers.

• Determine what combinations of products and service lines customers are likely to purchase and when.

• Analyze clinical trials for experimental drugs.

• Set more profitable rates for insurance premiums.

• Reduce equipment downtime by applying predictive maintenance.

• Determine with attrition and churn analysis why customers leave for competitors and/or become the customers.

• Detect and deter fraudulent behavior, such as from usage spikes when credit or phone cards are stolen.

• Identify promising new molecular drug compounds

Customers are the most critical aspect to a company's success. Without them a company cannot exist. So it is very important that firms have information on their preferences. Firms must quickly adapt to their changing demands. Business Intelligence enables firms to gather information on the trends in the marketplace and come up with innovative products or services in anticipation of customer's changing demands.

Competitors can be a huge hurdle on firm's way to success. Their objectives are the same as firms' and that is to maximize profits and customer satisfaction. In order to be successful firms must stay one step ahead of the competitors. In business, we don't want to play the catch-up

game because we would have lost valuable market share. Business Intelligence tells what actions our competitors are taking, so one can make better informed decisions. [3]

## 5. Business Intelligence Technology:

Business intelligence provides organizational data in such a way that the organizational knowledge filters can easily associate with this data and turn it into information for the organization. Persons involved in business intelligence processes may use application software and other technologies to gather, store, analyze, and provide access to data, and present that data in a simple, useful manner. The software aids in Business performance management, and aims to help people make "better" business decisions by making accurate, current, and relevant information available to them when they need it. Some businesses use data warehouses because they are a logical collection of information gathered from various operational databases for the purpose of creating business intelligence. [3]

In order for BI system to work effectively there must be some technical constraints in place. BI technical requirements have to address the following issues:

- Security and specified user access to the warehouse
- Data volume (capacity)
- How long data will be stored (data retention)
- Benchmark and performance targets

People working in business intelligence have developed tools that ease the work, especially when the intelligence task involves gathering and analyzing large quantities of unstructured data. Each vendor typically defines Business Intelligence their own way, and markets tools to do BI the way that they see it. Business intelligence includes tools in various categories, including the following:

- AQL - Associative Query Logic
- Score carding
- Business Performance Management and Performance Measurement
- Business Planning

## 6. Conclusion

Powerful transaction-oriented information systems are now commonplace in every major industry, effectively leveling the playing field for corporations around the world. To remain competitive, however, now requires analytically oriented systems that can revolutionize a company's ability to rediscover and utilize information they already own. The business intelligence (BI) has evolved over the past decade to rely increasingly on real time data. The BI systems auto-initiate actions to systems based on rules and context to support several business processes. These analytical systems derive insight from the wealth of data available, delivering information that's conclusive, factbased, and actionable. Enterprises today demand quick results. It is becoming essential nowadays that not only is the business analysis done, but also actions in response to analysis of results can be performed and instantaneously changes parameters of business processes. The paper explored the concepts of BI, its components, benefits of BI, technology requirements, designing and implementing business intelligence, and various BI techniques.

# Chapter 2

# Data Warehouse

## 1. What is Data warehouse

Different people have different definitions for a data warehouse which making it difficult to formulate a rigorous definition. Data warehouse are central repositories of integrated data from one or more disparate sources like databases (Oracle, SQL Server … etc.) flat files or other applications. They store current and historical data in one single place [4] and are used for creating analytical reports for knowledge workers throughout the enterprise. Examples of reports could range from annual and quarterly comparisons and trends to detailed daily sales analysis [4][5].

Data warehouse is a database designed to support decision making in organizations. It is updated in batches or in real time, and it is structured for rapid online queries and for providing managerial summaries. Data warehouses contain large amounts of historical data. According to William H . Inmon , who is one of earliest architectures they put it the construction of data warehouse systems "A data warehouse is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management's decision-making process." This definition is short but show and present the most major features of data warehouse.

**Subject oriented**: Data in an organization or company is organized in major objects or business processes. The common example of subject oriented data is customer, product, vendor and sale transaction.

**Integrated**: Data warehouse integrates data from various sources across departments within the organization. Different data sources can have different ways to define a specific object, for example, product or customers we can found customers in different sources CRM, financial system. However, in a data warehouse, there must be only one definition of products. This is achieved by using name conflict resolution in the data warehouse. And we when we achieve to make single version this we say the data is integrated.

**Non-volatile**: Data in the data warehouse is not subject to change . Once placed in a data warehouse , data is not updated.

**Time-variant**: Data in data warehouse associates with time. The time can be a single moment or span of time. In any cases, the data is only relevant to the time. [6]

## 2. Why Data warehouse

The separation of operational databases from data warehouses is based on the different structures, contents, and uses of the data in these two systems. Decision support requires historical data, whereas operational databases do not typically maintain historical data. In this context, the data in operational databases, though abundant, is usually far from complete for decision making. Decision support requires consolidation (such as aggregation and summarization) of data from heterogeneous sources, resulting in high-quality, clean, and integrated data. In contrast, operational databases contain only detailed raw data, such as transactions, which need to be consolidated before analysis. Because the two systems provide quite different functionalities and require different kinds of data, it is presently necessary to maintain separate databases. However, many vendors of operational relational database management systems are beginning to optimize such systems to support OLAP queries. As this trend continues, the separation between OLTP and OLAP systems is expected to decrease. [7]

## 3. Differences between Operational Database Systems and Data Warehouses [7]

Because most people are familiar with commercial relational database systems, it is easy to understand what a data warehouse is by comparing these two kinds of systems. The major task of on-line operational database systems is to perform on-line transaction and query processing. These systems are called on-line transaction processing (OLTP) systems. They cover most of the day-to-day operations of an organization, such as purchasing, inventory, manufacturing, banking, payroll, registration, and accounting. Data warehouse systems, on the other hand, serve users or knowledge workers in the role of data analysis and decision making. Such systems can organize and present data in various formats in order to accommodate the diverse needs of the different users. These systems are known as on-line analytical processing (OLAP) systems.

The major distinguishing features between OLTP and OLAP are summarized as follows:

### 3.1.Users and system orientation:

An OLTP system is customer-oriented and is used for transaction and query processing by clerks, clients, and information technology professionals. An OLAP system is market-

oriented and is used for data analysis by knowledge workers, including managers, executives, and analysts.

### 3.2.Data contents:

An OLTP system manages current data that, typically, are too detailed to be easily used for decision making. An OLAP system manages large amounts of historical data, provides facilities for summarization and aggregation, and stores and manages information at different levels of granularity. These features make the data easier to use in informed decision making.

### 3.3.Database design:

An OLTP system usually adopts an entity-relationship (ER) data model and an application-oriented database design. An OLAP system typically adopts either a star or snowflake model and a subject oriented database design.

### 3.4. View:

An OLTP system focuses mainly on the current data within an enterprise or department, without referring to historical data or data in different organizations. In contrast, an OLAP system often spans multiple versions of a database schema, due to the evolutionary process of an organization. OLAP systems also deal with information that originates from different organizations, integrating information from many data stores. Because of their huge volume, OLAP data are stored on multiple storage media.

### 3.5.Access patterns:

The access patterns of an OLTP system consist mainly of short, atomic transactions. Such a system requires concurrency control and recovery mechanisms. However, accesses to OLAP systems are mostly read-only operations (because most data warehouses store historical rather than up-to-date information), although many could be complex queries. Other features that distinguish between OLTP and OLAP systems include database size, frequency of operations, and performance metrics. These are summarized in Table 2.1). [7]

| *Feature* | *OLTP* | *OLAP* |
|---|---|---|
| Characteristic | operational processing | informational processing |
| Orientation | transaction | analysis |
| User | clerk, DBA, database professional | knowledge worker (e.g., manager, executive, analyst) |
| Function | day-to-day operations | long-term informational requirements, decision support |
| DB design | ER based, application-oriented | star/snowflake, subject-oriented |
| Data | current; guaranteed up-to-date | historical; accuracy maintained over time |
| Summarization | primitive, highly detailed | summarized, consolidated |
| View | detailed, flat relational | summarized, multidimensional |
| Unit of work | short, simple transaction | complex query |
| Access | read/write | mostly read |
| Focus | data in | information out |
| Operations | index/hash on primary key | lots of scans |
| Number of records accessed | tens | millions |
| Number of users | thousands | hundreds |
| DB size | 100 MB to GB | 100 GB to TB |
| Priority | high performance, high availability | high flexibility, end-user autonomy |
| Metric | transaction throughput | query throughput, response time |

*Tableau 1 Comparison between OLTP and OLAP systems [7]*

## 4. Data Warehouse Architecture

Data warehouse Architecture is a design that encapsulates all the facets of data warehousing for an enterprise environment. Data warehousing is the creation of a central domain to store complex, decentralized enterprise data in a logical unit that enables data mining, business intelligence, and overall access to all relevant data within an organization. Data warehouse architecture is inclusive of all reporting requirements, data management, security requirements, band width requirements, and storage requirements. There are three common types of data architecture which are as follows:

### 4.1.DATA WAREHOUSE ARCHITECTURE – BASIC:

Figure below shows a simple architecture for a data warehouse. End users directly access data derived from several source systems through the data warehouse. In this figure the metadata and raw data of traditional OLTP system is present, as is an additional type of data, summary data. Summaries are very valuable in data warehouses because they pre-compute long operations in advance. For example, a typical data warehouse query is to retrieval something like August sales.

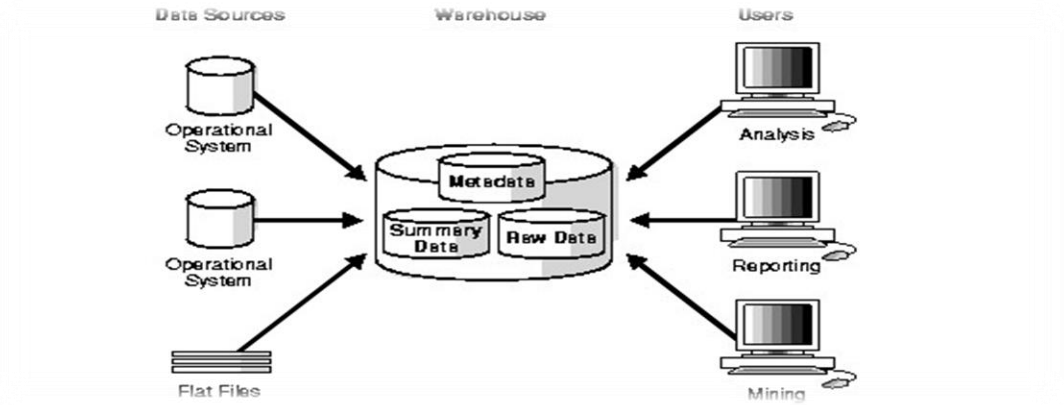



*Figure 2 Basic data warehouse architecture*

## 4.2. DATA WAREHOUSE ARCHITECTURE - WITH STAGING AREA

We need to clean and process the operational data before putting it into the warehouse. We can do this programmatically, although most data warehouses use a staging area instead. A staging area simplifies building summaries and general warehouse management.

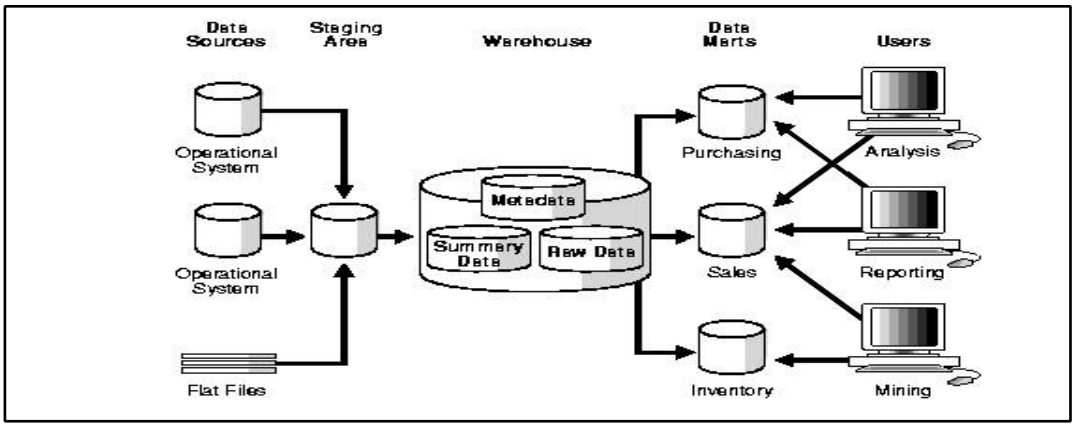



*Figure 3 staging data warehouse architecture*

### 4.3.Data Staging Area:

A place where data is processed before entering the warehouse. Data staging area is where the raw operational data is extracted, cleaned, transformed and combined so that it can be reported on and queried by users. This area lies between the operational source systems and the user database and is typically not accessible to users.

## 4.4.DATA WAREHOUSE ARCHITECTURE - WITH A STAGING AREA AND DATA MARTS

We may want to customize your warehouse's architecture for different groups within your organization. You can do this by adding data marts, which are designed for a particular line of business. An example is where purchasing, sales, and inventories are separated. In this example, a financial analyst might to analyze historical data for purchases and sales.

### 4.5. Data Mart:

Data mart is a logical subset of an enterprise-wide data warehouse. For example, a data warehouse for a retail chain is constructed incrementally from individual, conformed data marts dealing with separate subject areas such as product sales. Dimensional Data marts are organized

by subject area such as sales, finance and marketing and coordinated data category such as customer, product and location. These flexible information stores allows data structures to respond to business changes-product line additions, new staff, responsibilities, mergers, consolidations, and acquisitions[8].

# CHAPTER 3

# Data Mining and Enterprise Reporting

## 1. Data mining

**Data mining** is the process of discovering interesting knowledge, such as patterns, associations, changes, anomalies and significant structures, from large amounts of data stored in databases, data warehouses, or other information repositories. Due to the wide availability of huge amounts of data in electronic forms, and the imminent need for turning such data into useful information and knowledge for broad applications including market analysis, business management, and decision support, data mining has attracted a great deal of attention in information industry in recent years. [9]

Data mining has been popularly treated as a synonym of knowledge discovery in databases, although some researchers view data mining as an essential step of knowledge discovery. In general, a knowledge discovery process consists of an iterative sequence of the following steps:

- **Data cleaning**, which handles noisy, erroneous, missing, or irrelevant data
- **Data integration**, where multiple, heterogeneous data sources may be integrated into one.
- **Data selection**, data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operation.
- **Data mining**, which is an essential process where intelligent methods are applied in order to extract data patterns.
- **Pattern evaluation**, which is to identify the truly interesting patterns representing knowledge based on some interestingness measures.
- **Knowledge presentation**, where visualization and knowledge to the user.[9]

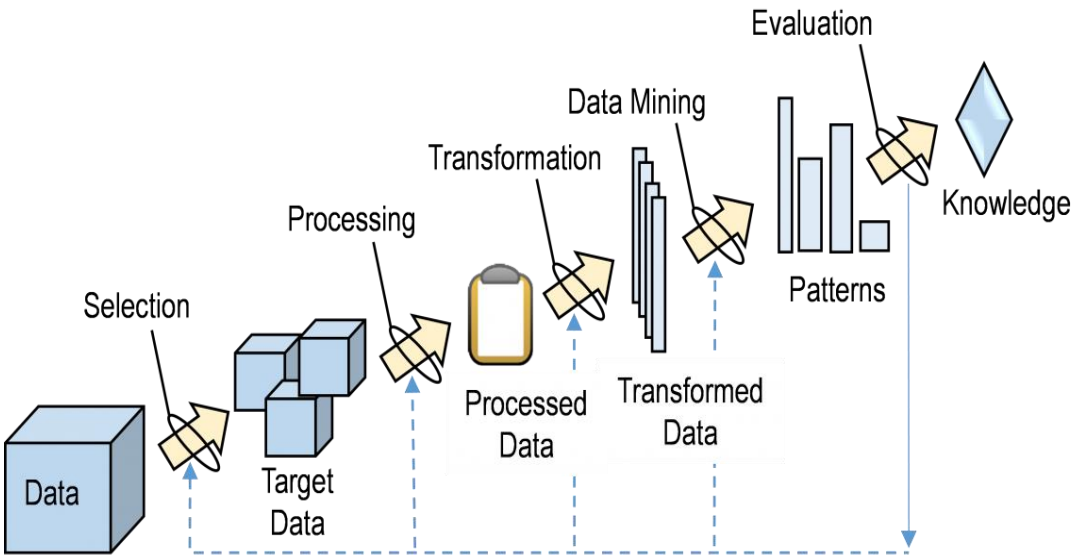



*Figure 4 data mining knowledge process.*

## 2. What kind of information are we collecting? [21]

We have been collecting a myriad of data, from simple numerical measurements and text documents, to more complex information such as spatial data, multimedia channels, and hypertext documents. Here is a non-exclusive list of a variety of information collected in digital form in databases and in flat files.

• **Business transactions**: Every transaction in the business industry is (often) "memorized" for perpetuity. Such transactions are usually time related and can be inter-business deals such as purchases, exchanges, banking, stock, etc., or intra-business operations such as management of in-house wares and assets. Large department stores, for example, thanks to the widespread use of bar codes, store millions of transactions daily representing often terabytes of data. Storage space is not the major problem, as the price of hard disks is continuously dropping, but the effective use of the data in a reasonable time frame for competitive decision-making is definitely the most important problem to solve for businesses that struggle to survive in a highly competitive world.

• **Scientific data:** Whether in a Swiss nuclear accelerator laboratory counting particles, in the Canadian forest studying readings from a grizzly bear radio collar, on a South Pole iceberg gathering data about oceanic activity, or in an American university investigating human psychology, our society is amassing colossal

amounts of scientific data that need to be analyzed. Unfortunately, we can capture and store more new data faster than we can analyze the old data already accumulated.

- **Medical and personal data**: From government census to personnel and customer files, very large collections of information are continuously gathered about individuals and groups. Governments, companies and organizations such as hospitals, are stockpiling very important quantities of personal data to help them manage human resources, better understand a market, or simply assist clientele. Regardless of the privacy issues this type of data often reveals, this information is collected, used and even shared. When correlated with other data this information can shed light on customer behavior and the like.

• **Surveillance video and pictures:** With the amazing collapse of video camera prices, video cameras are becoming ubiquitous. Video tapes from surveillance cameras are usually recycled and thus the content is lost. However, there is a tendency today to store the tapes and even digitize them for future use and analysis.

• **Satellite sensing**: There is a countless number of satellites around the globe: some are geo-stationary above a region, and some are orbiting around the Earth, but all are sending a non-stop stream of data to the surface. NASA, which controls a large number of satellites, receives more data every second than what all NASA researchers and engineers can cope with. Many satellite pictures and data are made public as soon as they are received in the hopes that other researchers can analyze them.

• **Games**: Our society is collecting a tremendous amount of data and statistics about games, players and athletes. From hockey scores, basketball passes and car-racing lapses, to swimming times, boxer's pushes and chess positions, all the data are stored. Commentators and journalists are using this information for reporting, but trainers and athletes would want to exploit this data to improve performance and better understand opponents.

• **Digital media:** The proliferation of cheap scanners, desktop video cameras and digital cameras is one of the causes of the explosion in digital media repositories. In addition, many radio stations, television channels and film studios are digitizing their audio and video collections to improve the management of their multimedia assets. Associations such as the NHL and the NBA have already started converting their huge game collection into digital forms.

• **CAD and Software engineering data:** There are a multitude of Computer Assisted Design (CAD) systems for architects to design buildings or engineers to conceive system

components or circuits. These systems are generating a tremendous amount of data. Moreover, software engineering is a source of considerable similar data with code, function libraries, objects, etc., which need powerful tools for management and maintenance.

• **Virtual Worlds**: There are many applications making use of three-dimensional virtual spaces. These spaces and the objects they contain are described with special languages such as VRML. Ideally, these virtual spaces are described in such a way that they can share objects and places. There is a remarkable amount of virtual reality object and space repositories available. Management of these repositories as well as content-based search and retrieval from these repositories are still research issues, while the size of the collections continues to grow.

• **Text reports and memos (e-mail messages):** Most of the communications within and between companies or research organizations or even private people, are based on reports and memos in textual forms often exchanged by e-mail. These messages are regularly stored in digital form for future use and reference creating formidable digital libraries.

• **The World Wide Web repositories:** Since the inception of the World Wide Web in 1993, documents of all sorts of formats, content and description have been collected and inter-connected with hyperlinks making it the largest repository of data ever built. Despite its dynamic and unstructured nature, its heterogeneous characteristic, and its very often redundancy and inconsistency, the World Wide Web is the most important data collection regularly used for reference because of the broad variety of topics covered and the infinite contributions of resources and publishers. Many believe that the World Wide Web will become the compilation of human knowledge

## 3. Major tasks of data mining [22]

In general, data mining tasks can be classified into two categories: descriptive data mining and predictive data mining . The former describes the data set in a concise and summary manner and presents interesting general properties of the data ; whereas the latter constructs one or a set of models , performs inference on the available set of data attempts to predict the behavior of new data sets .

A data mining system may accomplish one or more of the following data mining tasks

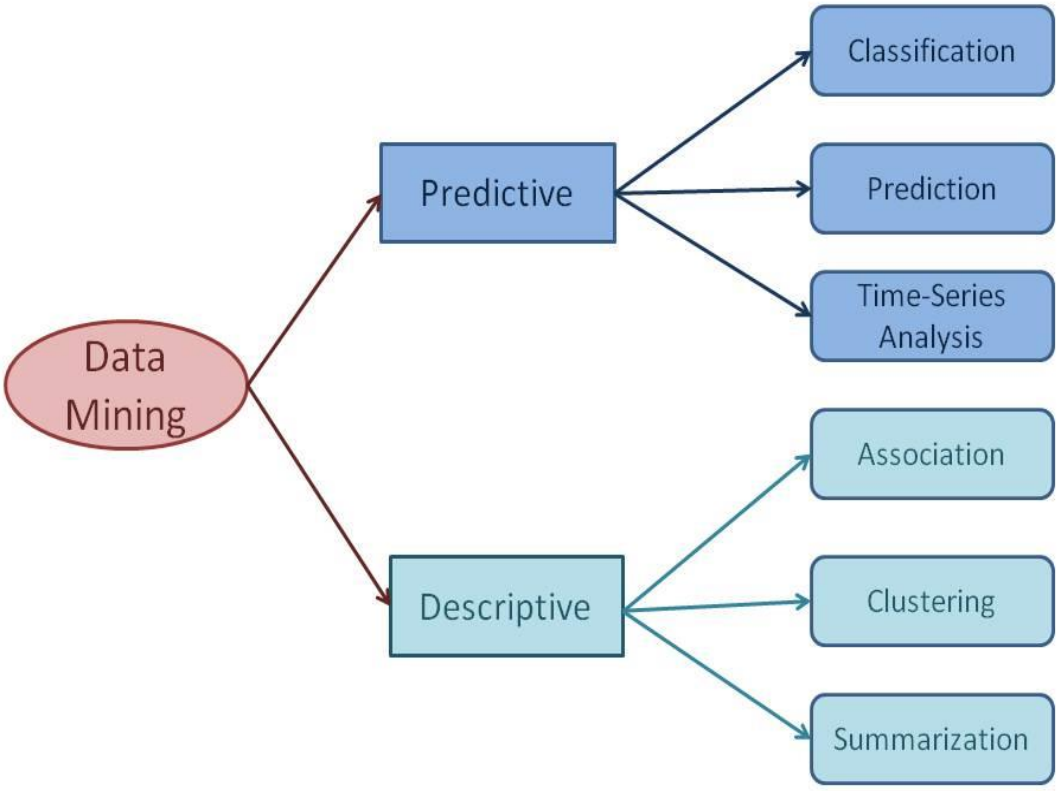



*Figure 5 data mining tasks*

### 3.1.Class description:

Class description provides a concise and succinct summarization of a collection of data and distinguishes it from others

The summarization of a collection of data is called class characterization; whereas the comparison or discrimination. Class description should cover not only its summary properties, such as count sum and average but also its properties on data dispersion such as variance quartiles, etc.

For example class description can be used to compare European versus Asian sales of a company , identify the important factors which discriminate the two classes , and present a summarized overview .

### 3.2.Association

Association is the discovery of association relationships or correlations among a set of items .They are often expressed in the rule form showing attribute-value conditions that occur frequently together in a given set of data . An association rule in the form of X=>Y is interpreted as ‘ database triples that satisfy X are likely to satisfy Y “ Association analysis is widely used in transaction data analysis for directed marketing , catalog design , and other business decision marking process .

Substantial research has been performed recently on association analysis with efficient algorithms proposed including the level wise apriority search , mining multiple-level , multi-dimensional associations , mining associations for numerical , categorical and interval data , meta-pattern directed or constraint – based mining , and mining correlations

### 3.3.Classification .

classification analyzes a set of training data (i.e ,a set of objects whose class label is known ) and constructs a model for each class based on the features in the data . A decision tree or a set of classification rules is generated by such a classification process , which can be used for better understanding of each class in the database and for classification of future data , For example , one may classify diseases and help predict the kind of diseases based on the symptoms of patients

There have been many classification methods developed in the fields of machine learning statistics , database , neural network rough sets and others . classification has been used in customer segmentation , business modeling and credit analysis

### 3.4.Prediction

this mining function predicts the possible values of some missing data or the value distribution of certain attributes in a set of objects . it involves the finding of the set of attributes relevant to selected object. for example , an employee’s potential salary can be predicted based on the salary distribution of similar employees in the company .Usually regression analysis generalized linear model correlation analysis and decision trees are useful tools in quality prediction Genetic algorithms and neural network models are also popularly used in prediction

### 3.5.Clustering.

Clustering analysis is to identify clusters embedded in the data , where a cluster is a collection of data objects that are 'similar' to one another . Similarity can be expressed by distance functions , specified by users or experts . A good clustering method produces high quality clusters to ensure that the inter-cluster similarity is low and the intra-cluster similarity is high for example one may cluster the houses in an area according to their house category , floor area and geographical locations .

Data mining research has been focused on high quality and scalable clustering methods for large databases and multidimensional data ware-houses

### 3.6.Times series analysis.

Time series analysis is to analyze large set of time series data to find certain regularities and interesting characteristics including search for similar sequence or subsequences, ad mining sequential patterns , periodicities , trends and deviation , For example in may predict the trend of the stock values for a company based on its stock history , business situation competitors performance and current market .

There are also other data mining tasks, such as outlier analysis, etc. Identification of new data mining tasks to make better use the collected data itself is an interesting research topic.

## 4. Data mining phases [20]

In the beginning of 90's, the growing of market interest in data mining shows signs of exploding into widespread uptake especially with absence of an effect methodology for extracting and discovering knowledge from the big amount of data coming from different sources or applications. this how the CRISP-DM (Cross Industry Standard Process for Data Mining) initiative was born.

CRISP-DM stands for cross-industry process for data mining. The CRISP-DM methodology provides a structured approach to planning a data mining project. It is a robust and well-proven methodology. This model consists of six phases intended as a cyclical process:

- **Business Understanding**: includes determining business objectives, assessing the current situation, establishing data mining goals, and developing a project plan.
- **Data Understanding:** Once business objectives and the project plan are established, data understanding considers data requirements. This step can include initial data collection,

data description, data exploration, and the verification of data quality. Data exploration such as viewing summary statistics (which includes the visual display of categorical variables) can occur at the end of this phase. Models such as cluster analysis can also be applied during this phase, with the intent of identifying patterns in the data.

- **Data Preparation** : Once the data resources available are identified, they need to be selected, cleaned, built into the form desired, and formatted. Data cleaning and data transformation in preparation of data modeling needs to occur in this phase. Data exploration at a greater depth can be applied during this phase, and additional models utilized, again providing the opportunity to see patterns based on business understanding.

- **Modeling:** Data mining software tools such as visualization (plotting data and establishing relationships) and cluster analysis (to identify which variables go well together) are useful for initial analysis. Tools such as generalized rule induction can develop initial association rules. Once greater data understanding is gained (often through pattern recognition triggered by viewing model output), more detailed models appropriate to the data type can be applied. The division of data into training and test sets is also needed for modeling.

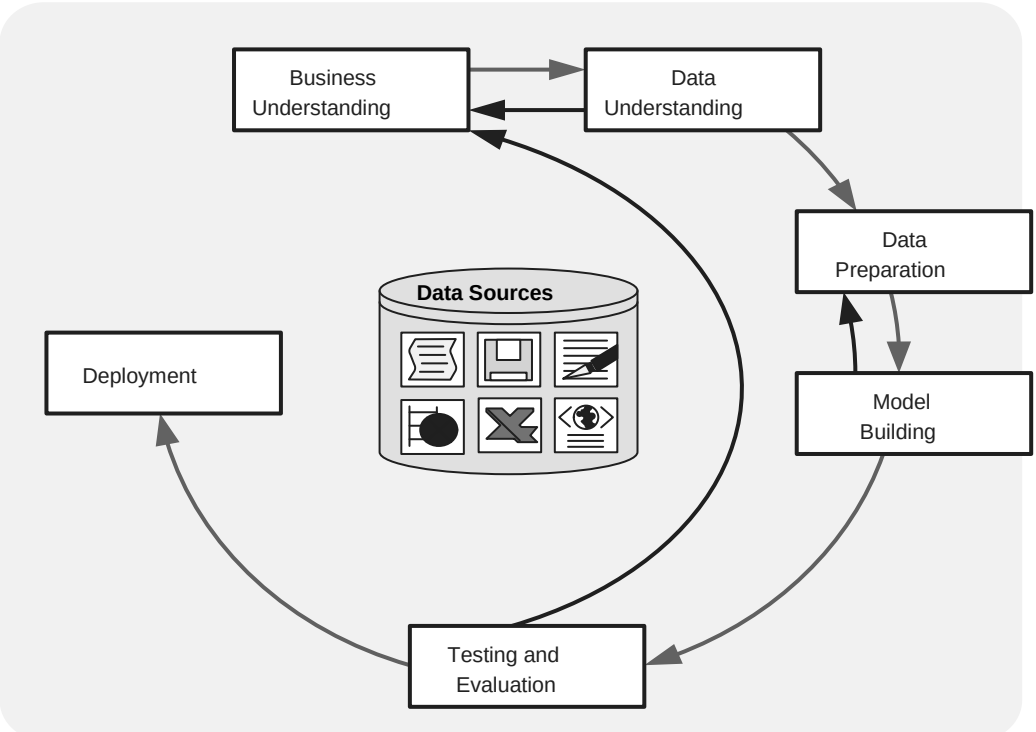



*Figure 6 CRISP-DM process [20]*

- **Evaluation Mode**l: results should be evaluated in the context of the business objectives established in the first phase (business understanding). This will lead to the identification of other needs (often through pattern recognition), frequently reverting to prior phases of CRISP-DM. Gaining business understanding is an iterative procedure in data mining, where the results of various visualization, statistical, and artificial intelligence tools show the user new relationships that provide a deeper understanding of organizational operations.
- **Deployment**: Data mining can be used to both verify previously held hypotheses, or for knowledge discovery (identification of unexpected and useful relationships). Through the knowledge discovered in the earlier phases of the CRISP-DM process, sound models can be obtained that may then be applied to business operations for many purposes, including prediction or identification of key situations. These models need to be monitored for changes in operating conditions, because what might be true today may not be true a year from now. If significant changes do occur, the model should be redone. It's also wise to record the results of data mining projects so documented evidence is available for future studies.

## 5. Data mining architecture [19]

Data mining is a very important process where potentially useful and previously unknown information is extracted from large volumes of data. There are a number of components involved in the data mining process. These components constitute the architecture of a data mining system.

The major components of any data mining system are data source, data warehouse server, data mining engine, pattern evaluation module, graphical user interface and knowledge base.

- **Data Sources**: Database, data warehouse, World Wide Web (WWW), text files and other documents are the actual sources of data. You need large volumes of historical data for data mining to be successful. Organizations usually store data in databases or data warehouses. Data warehouses may contain one or more databases, text files, spreadsheets or other kinds of information repositories. Sometimes, data may reside even in plain text files or spreadsheets. World Wide Web or the Internet is another big source of data.

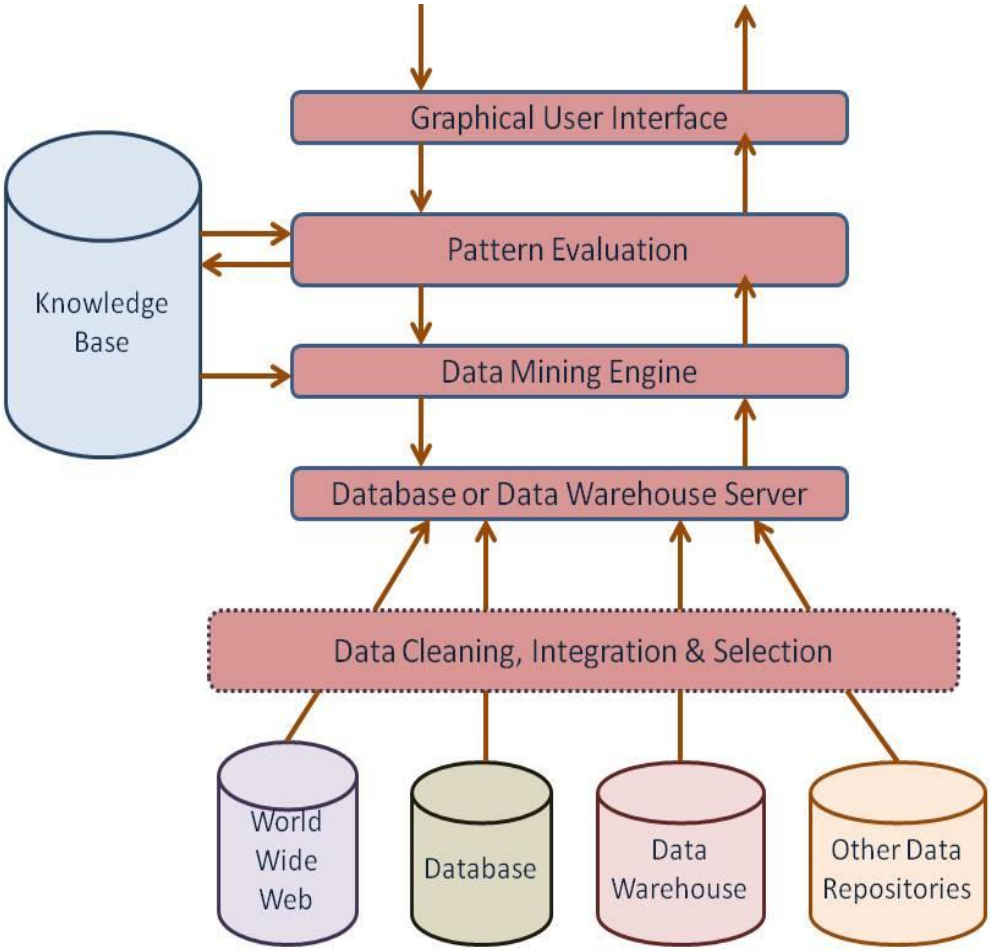



*Figure 7 Data Mining Architecture [20]*

- **Database or Data Warehouse Server:** The database or data warehouse server contains the actual data that is ready to be processed. Hence, the server is responsible for retrieving the relevant data based on the data mining request of the user.
- **Data Mining Engine**: The data mining engine is the core component of any data mining system. It consists of a number of modules for performing data mining tasks including association, classification, characterization, clustering, prediction, time-series analysis etc.
- **Pattern Evaluation Modules** The pattern evaluation module is mainly responsible for the measure of interestingness of the pattern by using a threshold value. It interacts with the data mining engine to focus the search towards interesting patterns.
- **Graphical User Interface** The graphical user interface module communicates between the user and the data mining system. This module helps the user use the system easily

and efficiently without knowing the real complexity behind the process. When the user specifies a query or a task, this module interacts with the data mining system and displays the result in an easily understandable manner.

- **Knowledge** Base The knowledge base is helpful in the whole data mining process. It might be useful for guiding the search or evaluating the interestingness of the result patterns. The knowledge base might even contain user beliefs and data from user experiences that can be useful in the process of data mining. The data mining engine might get inputs from the knowledge base to make the result more accurate and reliable. The pattern evaluation module interacts with the knowledge base on a regular basis to get inputs and also to update it.

Finally, each and every component of data mining system has its own role and importance in completing data mining efficiently. These different modules need to interact correctly with each other in order to complete the complex process of data mining successfully.

## 6. Data mining tools [18]

It is rightfully said that data is money in today's world. Along with the transition to an app-based world comes the exponential growth of data. But, most of data are unstructured and with the incredible growing of data that's make the process of extracting information and knowledge from data and transform into understandable and usable form to the decision-maker. These how data mining comes into the picture to help and introduce a many available tools for data mining tools using artificial intelligence and machine learning and other techniques of extracting data. here are most six famous tools available:

- **RapidMiner (formerly known as YALE) :** Written in the Java Programming language, this tool offers advanced analytics through template-based frameworks. A bonus: Users hardly have to write any code. Offered as a service, rather than a piece of local software, this tool holds top position on the list of data mining tools. In addition to data mining, RapidMiner also provides functionality like data preprocessing and visualization, predictive analytics and statistical modeling, evaluation, and deployment. What makes it even more powerful is that it provides learning schemes, models and algorithms from WEKA and R scripts.

RapidMiner is distributed under the AGPL open source licence and can be downloaded from SourceForge where it is rated the number one business analytics software.

- **WEKA :** The original non-Java version of WEKA primarily was developed for analyzing data from the agricultural domain. With the Java-based version, the tool is very sophisticated and used in many different applications including visualization and algorithms for data analysis and predictive modeling. Its free under the GNU General Public License, which is a big plus compared to RapidMiner, because users can customize it however they please. WEKA supports several standard data mining tasks, including data preprocessing, clustering, classification, regression, visualization and feature selection.

  WEKA would be more powerful with the addition of sequence modeling, which currently is not included.

- **R-Programming**: What if I tell you that Project R, a GNU project, is written in R itself? It's primarily written in C and Fortran. And a lot of its modules are written in R itself. It's a free software programming language and software environment for statistical computing and graphics. The R language is widely used among data miners for developing statistical software and data analysis. Ease of use and extensibility has raised R's popularity substantially in recent years.

Besides data mining it provides statistical and graphical techniques, including linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, and others.

- **Orange:** Python is picking up in popularity because it's simple and easy to learn yet powerful. Hence, when it comes to looking for a tool for your work and you are a Python developer, look no further than Orange, a Python-based, powerful and open source tool for both novices and experts.

You will fall in love with this tool's visual programming and Python scripting. It also has components for machine learning, add-ons for bioinformatics and text mining. It's packed with features for data analytics.

- **KNIME:** Data preprocessing has three main components: extraction, transformation and loading. KNIME does all three. It gives you a graphical user interface to allow for the assembly of nodes for data processing. It is an open source data analytics, reporting and integration platform. KNIME also integrates various components for machine

learning and data mining through its modular data pipelining concept and has caught the eye of business intelligence and financial data analysis.

Written in Java and based on Eclipse, KNIME is easy to extend and to add plugins. Additional functionalities can be added on the go. Plenty of data integration modules are already included in the core version.

- **Microsoft SQL Server Integration Service SSIS :** is platform for building enterprise level data integration and data transformations solutions . Integration Service is used to solve complex business problem by copying and downloading files, sending e-mail messages in response to events, updating data warehouses, cleaning and mining data, and managing SQL Server objects and data. The packages can work alone on in concert with other packages to address complex business needs. Integration Service can extract and transform data form a wide variety of sources such as XML data files, flat files, and relational data sources, and then load the data into one or more destinations.

## 7. Business Intelligence Reporting (BI Reporting)

BI reporting primarily enables in receiving output or results from a BI software solution. Typically, BI reporting is a preconfigured function, feature or component within a BI software solution. Based on the parameters set, BI reporting is generally an automated process that captures and report on analyzed data. These reports can be in form of statistical data, tables or visual charts and/or standard textual content.

The results content of BI reporting is generally in the form of actionable results that help the organization in short term, long term tactical and strategic decision making. It is also integrated with other application that takes the results data to perform any further operation process. [10]

### 7.1. BI Reporting Software

Picking the right reporting software for your business is a big step. It can unlock hidden value, help you make better decisions, and help you measure the impact of decisions you made in the past. [11] There’s many tools use it but most famous is:

- **Datadog Cloud Monitoring:** Datadog is the leading service for cloud-scale monitoring. It is used by IT, operations, and development teams who build and

operate applications that run on dynamic or high-scale infrastructure. Because Datadog collects metrics and events from 100+ different technologies and services out of the box, including other monitoring tools, you can monitor your entire stack in one place, without any gaps. With Datadog you can start monitoring in minutes without the need for ongoing maintenance**.**

- **IBM Cognos:** Cognos provides a cohesive Performance Management and Business Intelligence solution, with budgeting, strategic planning, forecasting, and consolidations - across financial, operational, sales and marketing, and human resources departments.
- **Report Builder 3.0:** is a report authoring environment for business users who prefer to work in the Microsoft Office environment. You work with one report at a time. You can modify a published report directly from a report server. You can quickly build a report by adding items from the Report Part Gallery provided by report designers from your organization.

## 8. Business Intelligence Dashboard

A business intelligence dashboard is a data visualization tool that displays the current status of metrics and key performance indicators (KPIs) for an enterprise. Dashboards consolidate and arrange numbers, metrics and sometimes performance scorecards on a single screen. They may be tailored for a specific role and display metrics targeted for a single point of view or department. The essential features of a BI dashboard product include a customizable interface and the ability to pull real-time data from multiple sources. [16]

### 8.1. BI Dashboard Software [17]

And just like the BI Reporting also BI Dashboard had many software free or paid we mentioned some off those software:

- **Agency Analytics:** Agency Analytics is an enterprise marketing dashboard that impresses clients and saves time. SEO tools like rank tracking, audits and backlinks are built in. Integrate other services like social media and PPC to get the whole picture. White label the dashboard and give clients their own personal logins to let them see your awesome marketing campaign at work.

- **iDashboards:** Headquartered in Troy, Mich., with customers in over 40 countries, iDashboards provides visually-appealing and cost-effective dashboard software for a wide variety of industries. With customizable dashboard solutions that easily integrate data sources, iDashboards clients can quickly and effectively analyze performance metricsleading to enhanced decision-making.
- **Klipfolio**: is a cloud app for building and sharing real-time business dashboards on web browsers, TV monitors and mobile devices. Connect to hundreds of data sources online and offline (including web applications like Twitter, Moz, Salesforce and Google Analytics), leverage a robust library of pre-built data visualizations, or build your own from scratch in a powerful Editor.

# CHAPTER 4

# CRM

In early of 90's business people started to use the term of CRM (Customers Relationship Management), when the concept of business started to change from being transactional to relational. CRM directly contributes towards customer benefits and the growth of businesses.

In these chapter, we will discuss about CRM basics and reporting CRM.

# 1. Definition

Customer Relationship Management (CRM) is a term representing the business strategy built around the concept of improved customer service. CRM practice involves all aspects of communication and dealing an organization has with its client, whether it is product or service linked. More clearly, CRM aims at increasing customer satisfaction, consequently increasing a business income. [11]

## 2. Objectives of CRM

In the commercial world the importance of retaining existing customers and expanding business is paramount. The costs associated with finding new customers mean that every existing customer could be important.

The more opportunities that a customer has to conduct business with your company the better, and one way of achieving this is by opening up channels such as direct sales, online sales, franchises, use of agents, etc. However, the more channels you have, the greater the need to manage your interaction with your customer base.

Customer relationship management (CRM) helps businesses to gain an insight into the behavior of their customers and modify their business operations to ensure that customers are served in the best possible way. In essence, CRM helps a business to recognize the value of its customers and to capitalize on improved customer relations. The better you understand your customers, the more responsive you can be to their needs. [12]

CRM can be achieved by:

- finding out about your customers' purchasing habits, opinions and preferences
- profiling individuals and groups to market more effectively and increase sales
- changing the way, you operate to improve customer service and marketing
- identifying which of your customers are profitable and which are not

## 3. Ingredients of CRM

CRM stands a

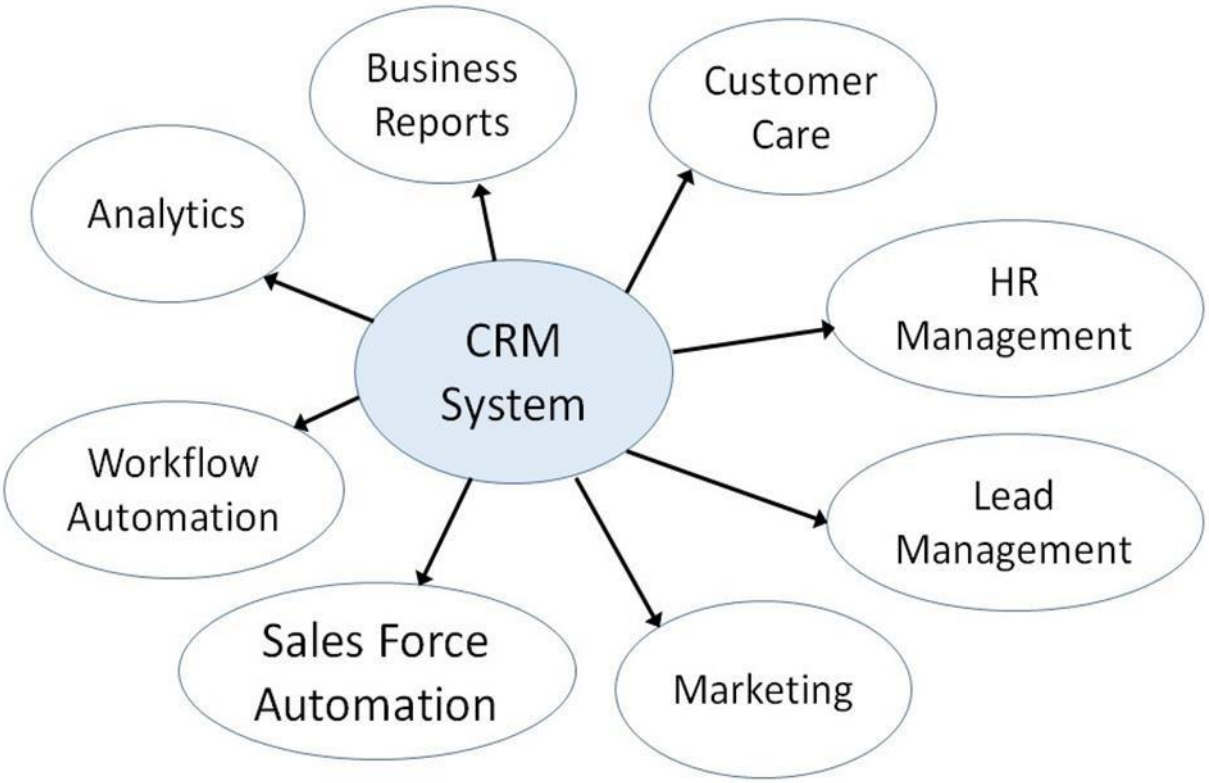



*Figure 8 Component of CRM*

- **Analytics**: Analytics is the process of studying, handling, and representing data in various graphical formats such as charts, tables, trends, etc., in order to observe market trends.

- **Business Reporting**: Business Reporting includes accurate reports of sales, customer care, and marketing.

- **Customer Service**: Customer Service involves collecting and sending the following customer-related information to the concerned department: o Personal information such as name, address, age, etc. o Previous purchase patterns. o Requirements and preferences. o Complaints and suggestions.

- **Human Resource Management**: Human Resource Management involves employing and placing the most eligible human resource at a required place in the business.

- **Lead Management:** Lead Management involves keeping a track of the sales leads and distribution, managing the campaigns, designing customized forms, finalizing the mailing lists, and studying the purchase patterns of the customers.

- **Marketing:** Marketing involves forming and implementing sales strategies by studying existing and potential customers in order to sell the product.

- **Sales Force Automation:** Sales Force Automation includes forecasting, recording sales, processing, and keeping a track of the potential interactions.

- **Workflow Automation:** Workflow Automation involves streamlining and scheduling various processes that run in parallel. It reduces costs and time, and prevents assigning the same task to multiple employees. [13]

## 4. Types of CRM

In the past twenty years, the focus of global markets has shifted from sellers to customers. Today, customers are more powerful than sellers, if we consider the driving factors of market. We have different types of CRM according to the changes in customer portfolios, speed of business operations, requirement of handling large data, and the need of sharing information, resources, and efforts jointly. [13]

CRM systems are divided based on their prominent characteristics. There are four basic types of CRM systems

- **Strategic CRM:** Strategic CRM is a type of CRM in which the business puts the customers first. It collects, segregates, and applies information about customers and market trends to come up with better value proposition for the customer.

  The business considers the customers' voice important for its survival. In contrast to Product-Centric CRM (where the business assumes customer requirements and focuses on developing the product that may sometimes lead to over-engineering), here the business constantly keeps learning about the customer requirements and adapting to them.

  These businesses know the buying behavior of the customer that happy customers buy more frequently than rest of the customers. If any business is not

considering this type of CRM, then it risks losing the market share to those businesses, which excel at strategic CRM. [13]

- **Operational CRM**: generally, refers to products and services that allow an organization to take care of their customers. It provides support for various business processes, which can include sales, marketing and service. Contact centers, data aggregation systems and web sites are a few examples. [14]
- **Analytical CRM:** is based on capturing, interpreting, segregating, storing, modifying, processing, and reporting customer-related data. It also contains internal business-wide data such as Sales Data (products, volume, purchasing history), Finance Data (purchase history, credit score) and Marketing Data (response to campaign figures, customer loyalty schemes data). Base CRM is an example of analytical CRM. It provides detailed analytics and customized reports. [13]
- **Collaborative CRM**: is an alignment of resources and strategies between separate businesses for identifying, acquiring, developing, retaining, and maintaining valuable customers. It is employed in B2B scenario, where multiple businesses can conduct product development, market research, and marketing. Collaborative CRM enables smooth communication and transactions among businesses. Though traditional ways such as air mail, telephone, and fax are used in communication, collaborative CRM employs new communication systems such as chat rooms, web forums, Voice over Internet Protocol (VoIP), and Electronic Data Interchange (EDI). [13]


*Figure 9 Collaborative CRM. [13]*

## 5. Data Warehouse for CRM [15]

A data warehouse is an information technology management tool that gives business decision makers instant access to information by collecting "islands of customer data" throughout the organization by combining all database and operational systems such as human resources, sales and transaction processing systems, financials, inventory, purchasing, and marketing systems. Specifically, data warehouses extract, clean, transform, and manage large volumes of data from multiple, heterogeneous systems, creating a historical record of all customer interactions (Eckerson and Watson, 2000). The abilities to view and manipulate set data warehouses apart from other computer systems. Constantly extracting knowledge about customers reduces the need for traditional marketing research tools such as customer surveys and focus groups. Thus, it is possible to identify and report by product or service, geographic region, distribution channel, customer group, and individual customer. Information is then available to all customer contact points in the organization. Data warehousing technology makes CRM possible because it consolidates, correlates and transforms customer data into customer intelligence that can used to form a better understanding of customer behavior. Customer data includes all sales, promotions, and customer service activities. In addition to transaction details, many other types of data generated from internal operations can make significant contributions. Information related to billing and account status, customer

service interactions, back orders, product shipment, product returns, claims history, and internal operating costs all can improve understanding of customers and their purchasing patterns. The ability of a data warehouse to store hundreds and thousands of gigabytes of data make drill-down analysis feasible as well as immediate. A corporate awareness survey conducted jointly by Cap Gemini and International Data Corporation (1999) found that 70 percent of US firms and 64 percent of European firms plan on building a data warehouse to support their CRM projects. A brief outline of organizational benefits with a data warehouse are:

- accurate and faster access to information to facilitate responses to customer questions.
- data quality and filtering to eliminate bad and duplicate0
- data extract, manipulate and drill-down data quickly for profitability analysis, customer profiling, and retention modeling.
- advanced data consolidation and data analysis tools for higher level summary as well as detailed reports,
- advanced data consolidation and data analysis tools for higher level summary as well as detailed reports;
- And calculate total present value and estimate future value of each and every customer.

## 6. Conclusion

Customer relationship management is a comprehensive approach that promises to maximize relationships with all customers, including Internet or “e-customers”, distribution channel members, and suppliers. Getting to “know” each customer through data mining techniques and a customer-centric business strategy helps the organization to proactively and consistently offer (and sell) more products and services for improved customer retention and loyalty over longer periods of time.

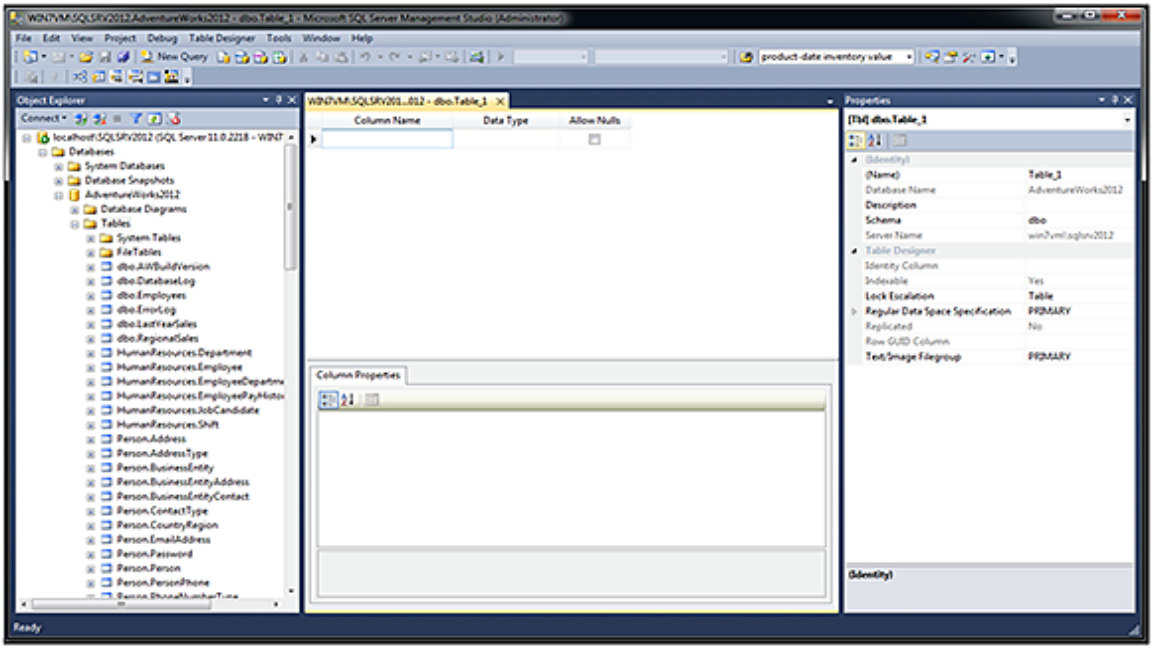
# Chapter 5

# Design and Implementation

# 1. Framework Tools

## 1.1.Microsoft SQL Server Entreprise Edition 2014 :

Microsoft SQL Server 2014 delivers several breakthrough capabilities that will enable your organization to scale database operation with confidence and Improve IT and developers efficiency, as well as enable highly scalable and well-managed business intelligence on a self-service basis for your users.



## 1.2.SQL Server Technologies:

SQL server management studio include many services to design data bases, to extract knowledge from those databases, and to show reporting based on this knowledge:

**1.2.1 The Database Engine** is the core service for storing, processing, and securing data. The Database Engine provides controlled access and rapid transaction processing to meet the requirements of the most demanding data consuming applications within your enterprise.

Use the Database Engine to create relational databases for online transaction processing or online analytical processing data. This includes creating tables for storing data, and database objects such as indexes, views, and stored procedures for viewing, managing,

and securing data. You can use SQL Server Management Studio to manage the database objects, and SQL Server Profiler to capture server events.

**1.2.2 Analysis Services** is an online analytical data engine used in decision support and business intelligence (BI) solutions, providing the analytical data for business reports and client applications such as Excel, Reporting Services reports, and other third-party BI tools. A typical workflow for Analysis Services includes building an OLAP or tabular data model, deploy the model as a database to an Analysis Services instance, process the database to load it with data, and then assign permissions to allow data access. When it's ready to go, this multi-purpose data model can be accessed by any client application supporting Analysis Services as a data source.

**1.2.3 Microsoft Integration** Services is a platform for building enterprise-level data integration and data transformations solutions. You use Integration Services to solve complex business problems by copying or downloading files, sending e-mail messages in response to events, updating data warehouses, cleaning and mining data, and managing SQL Server objects and data. The packages can work alone or in concert with other packages to address complex business needs. Integration Services can extract and transform data from a wide variety of sources such as XML data files, flat files, and relational data sources, and then load the data into one or more destinations.

**1.2.4 Integration Services** includes a rich set of built-in tasks and transformations; tools for constructing packages; and the Integration Services service for running and managing packages. You can use the graphical Integration Services tools to create solutions without writing a single line of code; or you can program the extensive Integration Services object model to create packages programmatically and code custom tasks and other package objects.

**1.2.5 Replication** is a set of technologies for copying and distributing data and database objects from one database to another and then synchronizing between databases to maintain consistency. Using replication, you can distribute data to different locations and to remote or mobile users over local and wide area networks, dial-up connections, wireless connections, and the Internet.

Transactional replication is typically used in server-to-server scenarios that require high throughput, including: improving scalability and availability; data warehousing and reporting; integrating data from multiple sites; integrating heterogeneous data; and offloading batch processing. Merge replication is primarily designed for mobile applications or distributed server applications that have possible

data conflicts. Common scenarios include: exchanging data with mobile users; consumer point of sale (POS) applications; and integration of data from multiple sites. Snapshot replication is used to provide the initial data set for transactional and merge replication; it can also be used when complete refreshes of data are appropriate. With these three types of replication, SQL Server provides a powerful and flexible system for synchronizing data across your enterprise.

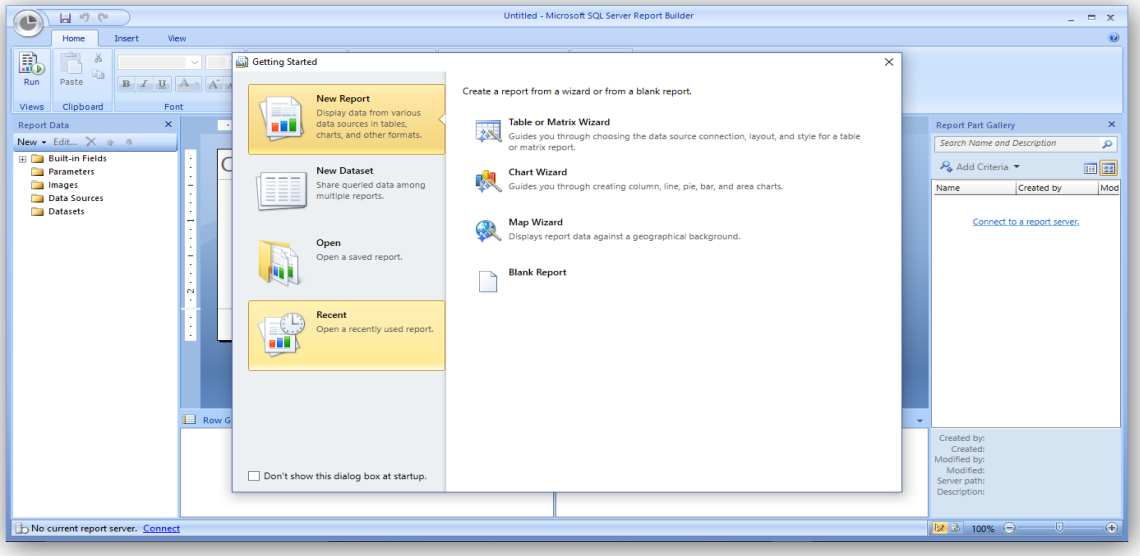
**1.2.6 SQL Server Reporting Services SSRS** provides a full range of ready-to-use tools and services to help you create, deploy, and manage reports for your organization. Reporting Services includes programming features that enable you to extend and customize your reporting functionality.

Reporting Services is a server-based reporting platform that provides comprehensive reporting functionality for a variety of data sources. Reporting Services includes a complete set of tools for you to create, manage, and deliver reports, and APIs that enable developers to integrate or extend data and report processing in custom applications. Reporting Services tools work within the Microsoft Visual Studio environment and are fully integrated with SQL Server tools and components.

With Reporting Services, you can create interactive, tabular, graphical, or free-form reports from relational, multidimensional, or XML-based data sources. Reports can include rich data visualization, including charts, maps, and sparklines. You can publish reports, schedule report processing, or access reports on-demand. You can select from a variety of viewing formats, export reports to other applications such as Microsoft Excel, and subscribe to published reports. The reports that you create can be viewed over a Web-based connection or as part of a Microsoft Windows application or SharePoint site. You can also create data alerts on reports published to a SharePoint site and receive email messages when report data changes.

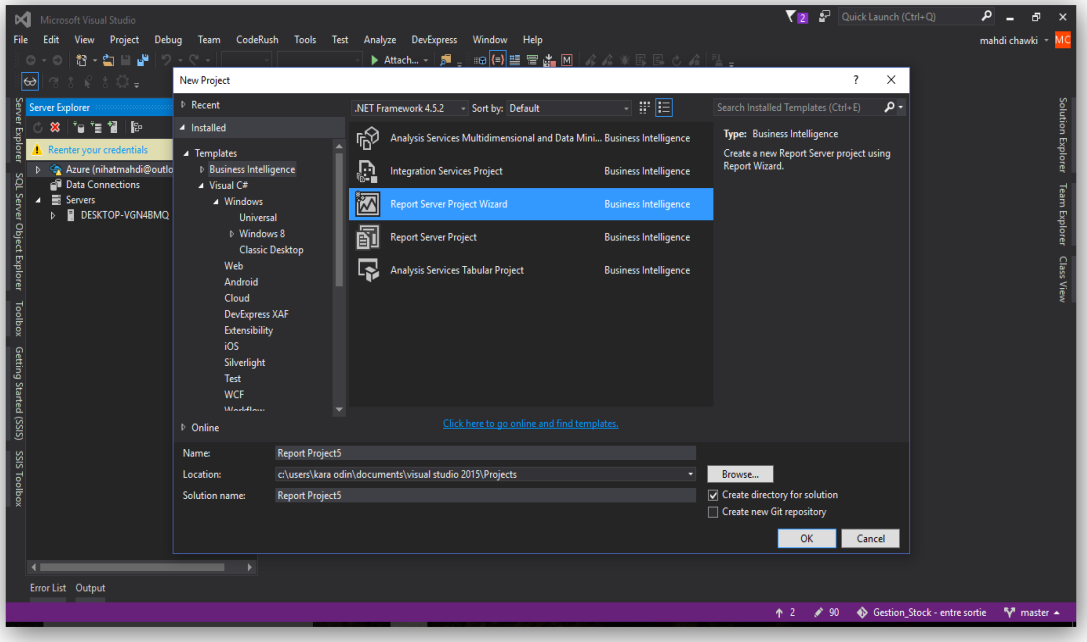
## 2. Microsoft SQL Server 2014 Report Builder 3.0

Report Builder provides data visualizations that include charts, maps, sparklines, and data bars that can help produce new insights well beyond what can be achieved with standard tables and charts. Use Report Builder to create reports and shared datasets. Publish report parts, and then browse the Report Part Gallery to reuse existing report parts as building blocks for creating new reports quickly with a “grab and go” experience.




## 3. SQL Server Data Tools SSDT 2015

SQL Server Data Tools (SSDT) transforms database development by introducing a ubiquitous, declarative model that spans all the phases of database development inside Visual Studio. You can use SSDT Transact-SQL design capabilities to build, debug, maintain, and refactor databases. You can work with a database project, or directly with a connected database instance on or off-premise.




In SSDT we have set of tools such as ETL and data mining tools:

- **SQL Server Integration Service SSIS:** In SQL Server Data Tools (SSDT), an Integration Services project stores and groups the files that are related to the package. For example, a project includes the files that are required to create a specific extract, transfer, and load (ETL) solution.

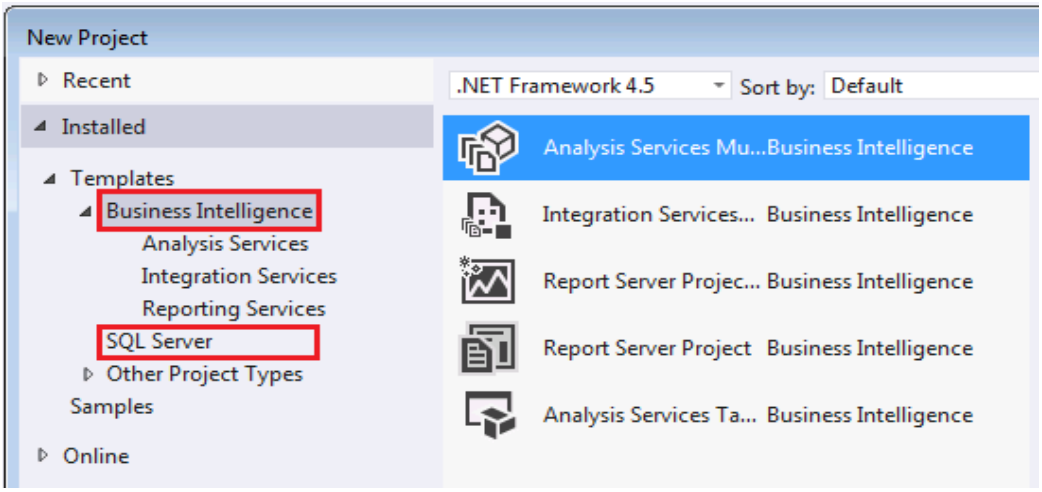
  When you add a new or an existing Integration Services project to a solution, SQL Server Data Tools (SSDT) creates project files that have the extensions .dtproj and .dtproj.user and .database.

  - The *.dtproj file contains information about project configurations and items such as packages.
  - The *.dtproj.user file contains information about your preferences for working with the project.
  - The *.database file contains information that SQL Server Data Tools (SSDT) requires to open the Integration Services project.
- **Reporting Services SSRS:** SQL Server Data Tools (SSDT) is a Microsoft Visual Studio environment for creating business intelligence solutions. SSDT features the Report Designer authoring environment, where you can open, modify, preview, save, and deploy Reporting Services paginated report definitions, shared data sources, shared datasets, and report parts. SQL Server Data Tools (SSDT) is not included with SQL Server.

  When you install Reporting Services, the following project templates are made available in SQL Server Data Tools (SSDT):

  - **Report Server Project:** When you select a Report Server Project, Report Designer opens. A Report Server Project is a Business Intelligence Projects template installed by Visual Studio that is available from the New Project dialog. Report Server project properties apply to all reports and all shared data sources in a SQL Server Data Tools (SSDT) project. These properties include the URL for the report server and the folder names for reports and shared data sources. Use the Project Property Pages dialog box to view the current property values. To open this dialog box, on the Project menu, click Properties.
  - **Report Server Project Wizard:** When you select a Report Server Wizard Project, a report server project is automatically created, and the

Report Wizard opens. In the wizard, you can create a report by following instructions on each page to create a connection string to a data source, set data source credentials, design a query, add a table or matrix data region, specify report data and groups, pick a font and color style, publish the report to a report server, and preview the report locally. After you create a report with the wizard, you can change the report data and the report designer by using Report Designer in the Report Server project



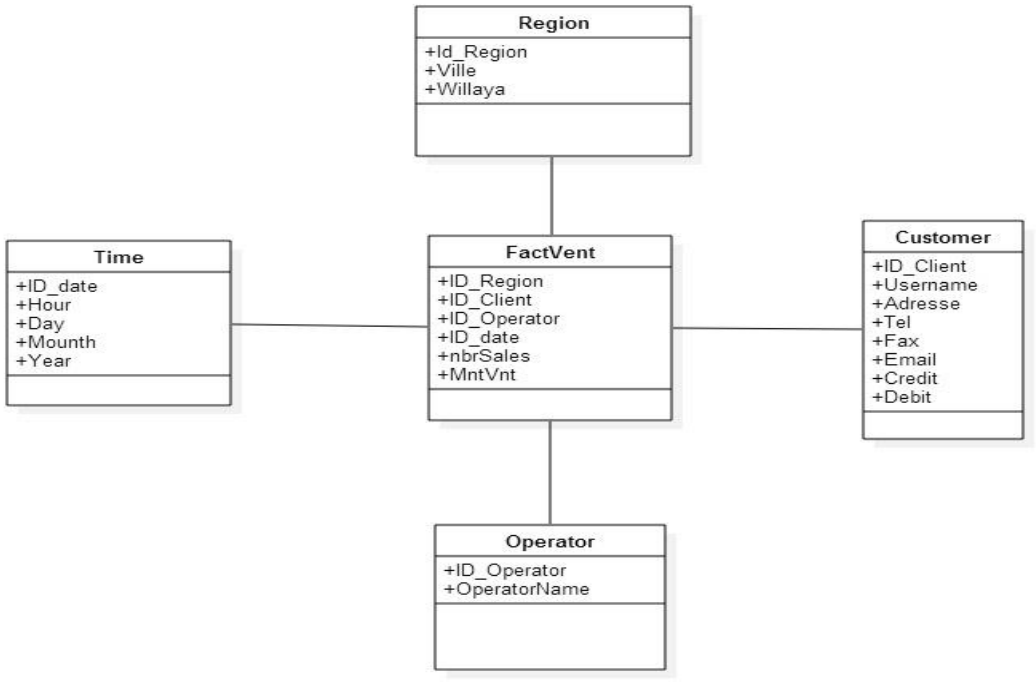
## 4. Implementation

Nessplus Telecom is an ISP and an enterprise for telecommunications and network solutions, they had many products such as Top-up and paid software for Medical solutions and e-marketing ..etc.

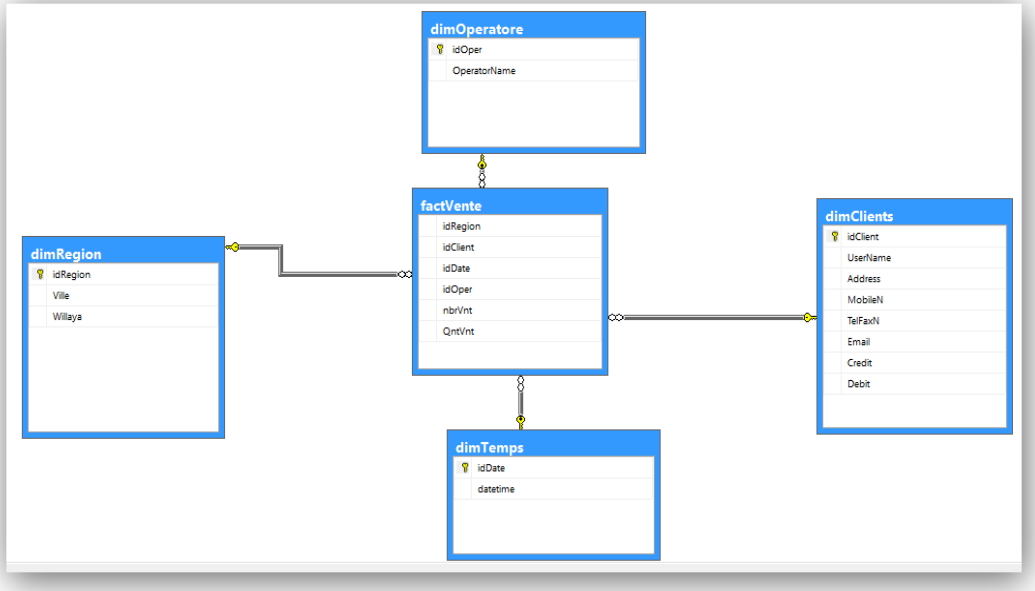
And like any enterprise in this world they have many customers and their database contain information for their activity especially with Top-up, in this section we are going to apply a CRM approach and methods to help decision-makers to know their customers situation and extract knowledge from a big database about historical records, to achieve that, first we are going to create data warehouse and with ETL SSIS we extracted and transformed data from a database into data warehouse, after that using SQL Server Report Builder we are going to render this knowledge in graphical form, tables or chart.

In first we are going to preparing ETL schema for extracting data to are data warehouse but before that, we are going to create a star schema for data warehouse in MS SQL Server 2014.

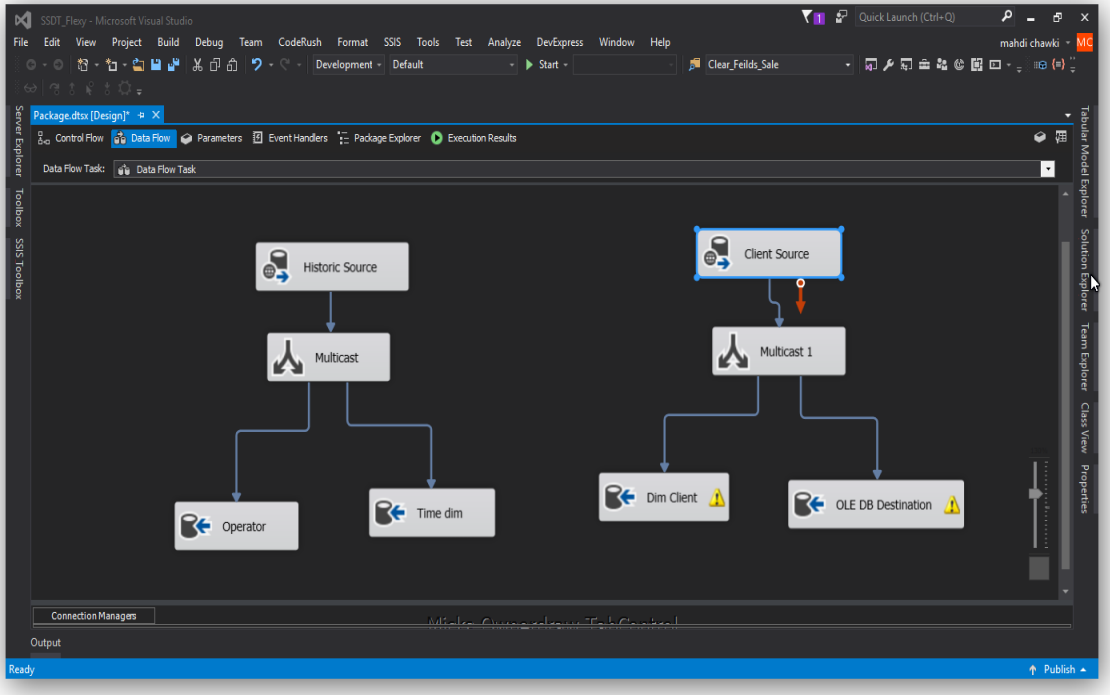
The fact table in are star schema is the FactVent because the main purpose of this data warehouse is to measure are sales by several dimensions like time geography and of course customers.



And here we have the diagram in SQL Server




And after, we create a data warehouse, now using SSIS we are going to create a ETL schema made us to extract and transform data into data warehouse from our data warehouse this, these sources can be from heterogenous system like our case MySQL system or from same system.



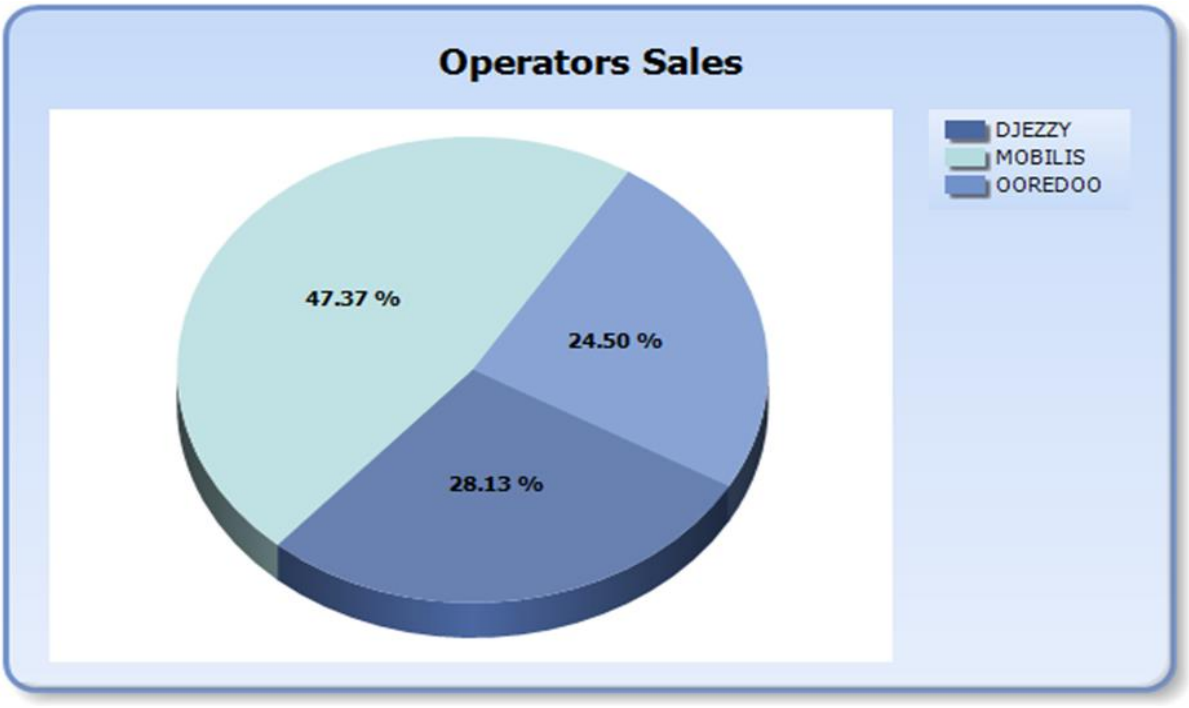
Until this moment, we create a data warehouse and ETL schema for extracting and transforming data into our container, and now using Report Builder we are going to generate a graphical form (chart, table), we are going to ask a questions usually asked by owners and we will see how we can get a useful knowledge for help them.

Q. -How much sales (number of sales) per customer, region and operator?

| Willaya | Client | DJEZZY | MOBILIS | OOREDOO | Total |
|---|---|---|---|---|---|
| Msila | NESSPLUS | 246485 | 2664085 | 299040 | 3209610 |
| | ABDEREZAK BOUTBIG | 320990 | 828290 | 513010 | 1662290 |
| | MOUNIR BERRA | 226310 | 625400 | 551260 | 1402970 |
| | BOULANOUAR LADJEL | 175560 | 559400 | 100420 | 835380 |
| | AYOUB LAKHDARI | 195200 | 498990 | 135410 | 829600 |
| | FOUAD BENYAHIA | 172590 | 433840 | 84550 | 690980 |
| | OUALID MOHAMED | 393200 | 137800 | 112550 | 643550 |
| | MOUNIR FARHAT | 149210 | 289820 | 83980 | 523010 |
| | ABDALLAH OUAAR | 122310 | 251455 | 56480 | 430245 |
| | AMINE SAHNOUNE | 112570 | 223190 | 61030 | 396790 |
| | RAOUF BENARIB | 72050 | 246465 | 60300 | 378815 |
| | HOUSSEM BARKAT | 144720 | 155560 | 63410 | 363690 |
| | HAMID BENYAHIA | 73370 | 227640 | 58680 | 359690 |
| | MOUNIM SLIMANI | 7090 | 8490 | 333660 | 349240 |
| | NESSPLUS 500 | 72780 | 164640 | 39320 | 276740 |
| | ABDEREZAK KERMICHE | 65880 | 171270 | 36650 | 273800 |
| | MOUSTFAOUI NOUREDDINE | 88150 | 138890 | 39410 | 266450 |
| | ISLAM LAKHAL | 43010 | 164210 | 27610 | 234830 |
| | KAISS BRIKI | 49600 | 125040 | 38180 | 212820 |
| | BILAL BENAMRA | 97960 | 52440 | 25500 | 175900 |
| | KAMEL ATIETALLAH | 23680 | 78190 | 67940 | 169810 |
| | BILAL HAMIDI | 27870 | 98030 | 19040 | 144940 |
| | ZEMOURI MOUNIR | 27930 | 82580 | 25130 | 135640 |
| | MOATAZ MAYOUF | 34160 | 62080 | 21050 | 117290 |
| | NABIL DOUMI | 23830 | 72600 | 15170 | 111600 |
| | BELHADAD TELECOM | 13350 | 18900 | 67060 | 99310 |
| | ABDEREZAK DJEGHDALI | 13540 | 69715 | 8880 | 92135 |

| | | | | | |
|---|---|---|---|---|---|
| | **MOHAMED KHERIFI** | 32020 | 47110 | 6430 | **85560** |
| | **ISMAIL OUALID** | 51240 | 24460 | 7320 | **83020** |
| | **MOHAMED SAOUDI** | 21011 | 29300 | 28370 | **78681** |
| | **MOHAMED MABKAR** | 19070 | 47440 | 6730 | **73240** |
| | **AMAR SLIMANI** | 11950 | 12010 | 35080 | **59040** |
| | **LAKHDAR HADIBI** | 11850 | 3740 | 3920 | **19510** |
| | **OMAR MAROUF** | 3400 | 8880 | 600 | **12880** |
| | **WEBSITE** | | 3000 | | **3000** |
| | **Total** | **3143936** | **8624950** | **3033170** | **14802056** |
| **Tlemcen** | **Total** | **2331655** | **739247** | **2294025** | **5364927** |
| **Souk Ahras** | **Total** | **656945** | **1929958** | **215450** | **2802353** |
| **Ghardaia** | **Total** | **286940** | **835845** | **980950** | **2103735** |
| **Batna** | **Total** | **303750** | **1493966** | **262150** | **2059866** |
| **Alger** | **Total** | **207750** | **136825** | **596240** | **940815** |
| **Guelma** | **Total** | **110540** | **201540** | **19080** | **331160** |
| **Laghouat** | **Total** | **80710** | **205445** | **33000** | **319155** |
| **Djelfa** | **Total** | **28030** | **145750** | **39520** | **213300** |
| **Oran** | **Total** | **45070** | **3750** | **19180** | **68000** |
| **Total** | | **11802063** | **19874512** | **10279775** | **41956350** |

Q. What is the percent of operators' sales?



# Conclusion

This works aims to introduce CRM as an effective method for maintain and improve the relationship between the customers and enterprises, in order to help this last to better understanding for their customers' needs and their mentality, depending on data warehouse and set of tools & scientific methods such as business reporting & data mining, for extracting useful knowledge help decision-makers to increase their economic gains and take better decision based on that extracted information.

In the mid-twentieth century, computerization invaded most of the economic fields which made the management of daily transaction more easy, like buy and selling. However, there is still a big problem, which is how to benefit from this big amount of rough data. With a well organization and better methods of extracting knowledge, CRM was supported by data warehouse which made a big leap from data transaction with large amount of rough and unrelated data to well-formed repository and methods for extracting and representing this knowledge in graphical form to the decision-makers, for design of a new business strategy faster and easier.

In this thesis, we take a live example for that Nessplus Telecom company. By taking the historical recording of their customers transaction in bad organized format and using SSDT tools such as SSIS and SSRS to extract and construct a data warehouse with star schema based on sales, and finally using a Report Builder, we represent this knowledge in graphical format (tables, chart. etc.), to give a better understanding and useful information about their customers situation.

## Future work

However, we must say that the experiments in this work are very preliminary, In the future work, we should design a kind of Cloud CRM offers access to the application via Web-based tools (or Web browser) logins where the CRM system administrator has previously defined access levels across the organization. Employees can log in to the CRM system, simultaneously, from any Internet-enabled computer or device. Often, cloud CRM provide users with mobile apps to make it easier to use the CRM on smartphones and tablets.

# References


[1] Dedić N. & Stanier C. (2016). Measuring the Success of Changes to Existing Business Intelligence Solutions to Improve Business Intelligence Reporting. Lecture Notes in Business Information Processing. Springer International Publishing. Volume 268, pp. 225-236.

[2] *What is Business Intelligence* Available : http://www.selectbs.com/products-general/what-is-business-intelligence [accessed 07-04-2017].

[3] JAYANTHI RANJAN "BUSINESS INTELLIGENCE: CONCEPTS, COMPONENTS, TECHNIQUES AND BENEFITS". *Journal of Theoretical and Applied Information Technology*: Vol 9. No 1. (pp 060 - 070).

[4] Dedić, N. and Stanier C., 2016., "An Evaluation of the Challenges of Multilingualism in Data Warehouse Development" in 18th International Conference on Enterprise Information Systems - ICEIS 2016, p. 196.

[5] *9 Reasons Data Warehouse Projects Fail* Available : https://blog.rjmetrics.com/2014/12/04/10-common-mistakes-when-building-a-data-warehouse/.[accessed 14-04-2017]

[6] *Data Warehouse Definition* Available : http://www.1keydata.com/datawarehousing/data-warehouse-definition.html [accessed 25-04-2017]

[7] *Jiawei H & Micheline K, Data Mining: Concepts and Techniques Second Edition*

[8] DATA WAREHOUSE ARCHITCTURE AND DATA ANALYSIS TECHNIQUES Mrs. Vaishali V. Khachane Dept of Information Technology VPM's Polytechnic Thane, Mumbai

[9] S.Chen, J.Hen, and P. S. Yu. Data Mining: An Overview from a database perspective. IEEE Transactions on Knowledge and Data Engineering, 8(6):866-883, 1996.

[10] *Business Intelligence Reporting (BI Reporting)* available in https://www.techopedia.com/definition/30217/business-intelligence-reporting-bi-reporting [ accessed in 20/05/2017]

[11] *CRM DEFINITION* http://www.esalestrack.com/crm_article/crm-definition.html [Accessed in: 16/05/2017]

[12] *CUSTOMER RELATIONSHIP MANAGEMENT* http://www.infoentrepreneurs.org/en/guides/customer-relationship-management/ [Accessed in: 16/05/2017]

[13] *CRM – Types* https://www.tutorialspoint.com/customer_relationship_management/crm_types.htm [Accessed in: 26/05/2017]

[14] *Customer Relationship Management (CRM)* http://www.straightmarketing.com/customer_relations_management.asp [Accessed in: 26/05/2017]

[15] J. Chen and Karen Popovich, Understanding customer relationship management (CRM) People, process and technology, Department of Operations Management and Business Statistics, College of Business Administration, Cleveland State University, Cleveland, Ohio, USA

[16] *business intelligence dashboard* hhttp://searchbusinessanalytics.techtarget.com/definition/business-intelligence-dashboard [Accessed in: 16/05/2017]

[17] *Dashboard Software* http://www.softwareadvice.com/bi/dashboard-comparison/ [Accessed in: 16/05/2017]

[18] *Six of the Best Open Source Data Mining Tools* https://thenewstack.io/six-of-the-best-open-source-data-mining-tools/ [Accessed in: 06/05/2017]

[19] *Data Mining Architecture* http://www.wideskills.com/data-mining-tutorial/data-mining-architecture [Accessed in: 06/05/2017]

[20] *What is the CRISP-DM methodology?* http://www.sv-europe.com/crisp-dm-methodology/ [Accessed in: 06/05/2017]

[21] *Chapter I: Introduction to Data Mining* https://webdocs.cs.ualberta.ca/~zaiane/courses/cmput690/notes/Chapter1/ [Accessed in: 05/05/2017]

[22] *Chengqi Zhang, Shichao Zhang, Association Rule Mining: Models and Algorithms*

**ملخص**. في هذا العالم المتغير ومع النمو الهائل للأسواق والمؤسسات في زمن العولمة، ومع المنافسة الكبيرة، والمحمومة لكسب المزيد من الأرباح ونيل ولاء ورضى العملاء. إدارة علاقة العملاء تدعمها التكنولوجيات، واستخراج البيانات. الخ. تمثل حل قوي وفعال لتحسين علاقات تجارية مع العملاء. في هذه الأطروحة، نحن نأخذ مثالا حيا لشركة Nessplus Telecom ، و نحاول تطبيق نهج إدارة علاقات العملاء من خلال إنشاء مستودع البيانات تحتوي على جميع المعاملات اليومية في شكل منظم، ونبين كيف يمكننا استخراج المعلومات المفيدة في شكل رسوم بيانية (الرسم البياني والجداول. )، بغية مساعدتهم على زيادة أرباحهم.

**كلمات مفتاحية:** ذكاء الأعمال، إدارة علاقة العملاء، تنقيب البيانات، الارباح، مستودع البيانات

**Abstract**. In this changing world and with the incredible growing of markets & enterprises during the globalization age, and with the big and strong competition to gain more profit and the customer loyalty and his satisfaction. CRM supported by technologies, data mining .etc. represented as an effectiveness and powerful solution for improve business relationships with customers. In this thesis, we take a live example for that Nessplus Telecom company, and we will to apply CRM approach by creating a data warehouse contain all daily transaction in well-formed, and show how we can extract a useful knowledge in graphic form (chart, tables...), for helps them to increase their gains.

**Key word**: BI, CRM, Data mining, Gains, Data warehouse.

**Résumé**. Dans ce monde changeant et avec la croissance incroyable des marchés et des entreprises durant de l'ère de la mondialisation d'une part, et avec la grande et la forte concurrence pour gagner plus de profit et de fidélité de la clientèle et pour sa satisfaction d'autre part, CRM a pris en charge l'exploration de données par plusieurs technologies. Ces données se représentent comme une puissante et une efficace solution pour améliorer les relations d'affaires avec les clients Dans cette thèse, on a pris un exemple direct pour cette société c'est : Nessplus Telecom, où on a appliqué une approche CRM en créant un entrepôt de données contenant toutes les transactions quotidiennes bien formées en montrant comment extraire une connaissance utile sous forme graphique (graphique, Tableaux ...) pour les aider à augmenter leurs gains.

**Mot clé** : BI, CRM, Exploration de données, Gains, Entrepôt de données.