POLITECNICO DI TORINO

Faculty of Engineering Engineering and Management

Master Degree Thesis

Implementation of a Business Intelligence solution in the Large-Scale Retail Trade Domain. The case of Iper.



Supervisor prof. Claudio Demartini Candidate Ginna Roxana GIRALDO OVALLE

October 2018

Dedication and acknowledgements

This thesis and my academic achievements are dedicated to God and to my beloved family, specially my mom Marcela Ovalle and my sister Yihan Maki, who have always been a constant source of support and encouragement during the challenges of my whole life, Thank you both for giving me strength to reach for the stars and chase my dreams.

I would like to thank the company FINIPER, for allowing me to be part of this process and to work on this project. In particular, I would like to direct my special thanks to my company supervisor Deris Gil Martinez, who provided me with tools and managerial knowledge. My special gratitude to Marco Saviano, Big Data Specialist who provided me all the tools and technical notions required to apply this thesis. Thank you for your guidance and support throughout this study.

I am grateful also to my supervisor Claudio Demartini, for accompanying me during the development of this thesis, your patience and valuable advice during the whole work.

Finally, I would like to express my gratitude to the Polytechnic of Turin and the Pontificia Universidad Javeriana, for the opportunity to bring cultures closer with the double degree program.

Table of Contents

			Pa	age
Li	st of	Figures		ix
1	Intr	roduction		1
2	Data	a Driven Organization		3
	2.1	Data-Driven Organization		3
	2.2	Data Collection		3
	2.3	Reporting		7
		2.3.1 Metric Design		8
	2.4	Analysis		10
		2.4.1 Type of analysis		11
		2.4.2 Levels of Analytics		14
		2.4.3 Analyst human component		16
	2.5	Analytics Organization Structure		18
		2.5.1 Centralized		19
		2.5.2 Decentralized		20
		2.5.3 Hybrid Model		20
	2.6	Decision Making		21
3	The	eoretical Framework		23
	3.1	System Architecture		23
		3.1.1 Data Transmission		25
	3.2	Analytics Maturity		25
		3.2.1 Technical Approach		27
		3.2.2 Non Technical Approach		30
		3.2.3 Information Systems		31
	3.3	Business Intelligence		33
		3.3.1 Reporting		35
		3.3.2 ETL and ELT		38

		3.3.3	BI Components: OLAP,	39
		3.3.4	BI Components: Data Discovery	39
4	Fini	iper Gr	oup	41
	4.1	Organ	izational Description	42
		4.1.1	About the company	42
		4.1.2	Mission	44
		4.1.3	Market focus	45
	4.2	Produ	cts	45
	4.3	Inform	nation Technology Department	47
5	As.T	s Arch	itecture	49
J	5 1	Archit	acture As-Is	49 49
	0.1	5 1 1		52
		510		55
		0.1.2		
		5.1.3	1 wittiper	57
		5.1.4	Microstrategy	60
	5.2	Organ		62
		5.2.1	Analytics Maturity	64
		5.2.2	Criticalities	65
6	To E	Be: Qlil	rsense	67
	6.1	Overv	iew	67
	6.2	Get cle	ose to the application	69
		6.2.1	Main Characteristics	69
7	Mat	hadala	and Implementation Disp	79
4	Met		By and implementation Fian.	70
	7.1 7.0	A9-19		13
	1.2	Implei		75
		7.2.1		76
		7.2.2	Organizational Structure	95
8	Con	clusio	as	101
	8.1	Gener	al conclusion	102
	8.2	Recom	mendations	103
A	Ann	endiv	Α	105
	А 1	Regist	 rv Table Product	105
	Δ9	Ragist	ry Table Providers	115
	л.2 Л 9	Domint	1y Table 1 Toviders	104
	п.э	regist	1 y Tanie Dales Divites	124

bliog	bliography 135						
A.8	Facts Table Accounting Stock	131					
A.7	Registry Table Merchandise Structure	130					
A.6	Fact Table Daily Sales	129					
A.5	Fact Table Assortment PDV	126					
A.4	Registry Table Calendar	125					

Bibliography

List of Figures

FIGURE

Page

2.1	Example. Importance of context	6
2.2	Iper Montebello Data Dictionary	7
2.3	The Analytics value Chain	7
2.4	Illustration of the metrics within a company	9
2.5	Subset of common KPI's.	10
2.6	Generalized linear model, Anova	13
2.7	Flexsim software simulation process.	14
2.8	Analytics Levels	15
2.9	Intersection between the level of analytics and type of analysis	16
2.10	Health, Case study	19
3.1	Spaghetti Structure	24
3.2	Organized Structure i	24
3.3	Organization Maturity	26
3.4	Maturity Stages of DWH	28
3.5	Relational Model	30
3.6	Example of SQL syntax	30
3.7	Gartner, Analytics Maturity	31
3.8	Abax Bi-Metrix ,Example of report.	36
3.9	Choosing a good visualization	37
3.10	Example of StoryTelling	38
4.1	Company control agents	42
4.2	Store Map and Markets, November 2017	43
4.3	Holding Organization Chart	44
4.4	Iper la grande, different brands	46
4.5	Iper la grande i,merchandise structure	46
4.6	Location Iper	47
4.7	Architecture E-commerce DSI	48

5.1	Architecture DSI	50
5.2	Structure Table DimForn	52
5.3	Cercacodice Main Characteristics	54
5.4	Scale of the survey applied	54
5.5	Level of utilization CercaCodice	54
5.6	Cercacodice Application	55
5.7	Webgate Main Characteristics	56
5.8	Webgate Level of utilization	56
5.9	Webgate Application	57
5.10	Twittiper Main Characteristics	59
5.11	Twittiper Level of utilization	59
5.12	Twittiper Application	60
5.13	Microstrategy Main Characteristics	61
5.14	Microstrategy Level of utilization	61
5.15	MicroStrategy Prompt details	62
5.16	Team Organization	63
5.17	Example Analysis Excel	64
5.18	Analytics Maturity	65
6.1	Enterprise Deployment	68
6.2	Qliksense Hub	69
6.3	Qliksense Create an Application	70
7.1	Overall Project Plan	75
7.2	Memory occupied by the QVD files.	78
7.3	Application introduction page	91
7.4	Application. Standard Filters	92
7.5	Application. Variable Filters	93
7.6	Application. Different pages	94
7.7	Metrics defined for the App \ldots	95
7.8	Initial proposal Analytics Team	97
7.9	Process proposed to be implemented	98

Chapter 1

Introduction

"If you can not measure it, you can not improve it." with this quote made by Peter Drucker [12], is intend to start this dissertation where the objective is to establish the importance of the data and how the processes of collection, management and interpretation are fundamental to drive the company to success through the decision-making process . All this is possible through the use of technology that has helped improving quality, efficiency and effectiveness within companies. During all these years, wonderful ideas have been growing and this is also the case of business intelligence tools, which allow to take advantage of all the information presented about the events that occurred during the services provided by the company, to anticipate the questions that were not already planned and analyze the reports of the business processes of a company in real time.

This thesis presents the analysis and implementation of the activities carried out during the support given to the IT department of the Italian company Iper Montebello, for the development of a implementation plan of a Business Intelligence solution, for the analytics processes carried out within the different departments. It is included also, the analysis of the tools that the company had decided to use, and the actual implementation of the solution Qliksense involving also the definition and creation of new procedures by changing the organizational structure in this field.

The Qliksense Business Intelligence tool offers a solution for *Analytics processes*, to the end users of different departments of the Italian retail company dedicated to offer services to the Italian consumers in the GDO field. Iper Montebello has been growing in the last years and it is the pioneer of the concept of hypermarket in Italy, following the principles by which initially was created, the company has evolve in order to empower the customers experience. The management of its analytics processes has been carried out in the same way during many years, time by which the emerging companies had opt by implementing and integrating the new technologies. Now, it is time to be concordant with the change and to align the strategy and business process in order to maintain a competitive advantage , analyzing those technologies and taking advantage of them, to be able to highlight in the competitive field of the large retail market. This has brought to the position of implementing Qliksense , over the tree solutions CercaCodice, Webgate and Twittiper that the company had been using.

The recognizing of concepts took place, for the objective of understanding the framework behind the analysis so, in this way there was included the *Data-Driven* culture and therefore, the Data collection processes were studied, the Business *Analytics Processes* for the actual definition of each of the processes, as well as the basic concepts of the *Data-Discovery* and then for sure introduce all the functionalities of the solution proposed *Qliksense*, including its components and main characteristics.

The first step on the Implementation is to evaluate Iper's *As-Is situation*, this means that it is necessary to find the characteristics that made of CercaCodice, WebGate and twittiper ,to be selected when the company designed the processes at the first moment, and to find the perimeter uncovered to make sure the new proposal provides a solution to each of them. Moreover, for each one of the process definitions, it has been carried out an analysis, making sure that requirements are covered and handled as expected . During this part of the analysis were involved the end users from the departments so the positive user experience was ensure with both perspectives, the business and technical points of view.

Similarly, the Qliksense solution is analyzed, having into account the aspects that were to change after the previous analysis. This is, to create what is presented as the *to-be situation*. It is also illustrated, the new organization of the system, in which the actual logic of the architecture is been changed and also the organizational structure.

As a conclusion, it can be highlighted the role of Qliksense as an important aspect of improvement for the company, because it defines standards , helps to increase the data governance in addition to provide worth trust and reliable data. The advantages on the solution corresponds with the flaws found with the previous implementation, which involved the CercaCodice, Web-Gate and twittiper. highlighting that Qliksense ensures different future technological updates regarding the analytics solutions. The presented document contributes to the understanding of implementation of new technologies in order to increment the data-driven culture within Iper organization. It is grateful to specify that the redefined processes are currently being deployed for one of the most critical departments, with high and positive replies and expectations.

Chapter 2

Data Driven Organization

What it is intended to be done in this dissertation, is to support the Data Intelligence activities of the Iper Montebello Company, which are in charge of the DSI Department, by making use of some engineering concepts and more in detail, the goal is to change the organizational culture in order to create a data-driven organization so that the information systems and tools generated and delivered by the department can be leveraged optimally and represent monetary benefits. To do this, it is necessary to put into context all the fundamentals and link them altogether.

2.1 Data-Driven Organization

A Data-Driven organizations is a concept that has been misunderstood among the years and has been seen as the processes that involves background-view or better explained, the company's consider that been a data driven company only includes the fact of gathering data and making some descriptive analysis about the facts, but a data driven organization is a matter of culture, culture in the way the processes of managing data are done, the idea is to consider a company as a machine, and the competences are measure in the way the resources or change agents are able to follow the customer requirements and the market adaptability through the different tools available in the market in order to do forward analysis trying to interpret and predict future scenarios that gives to the company "Machine" a competitive advantage making use of the right, relevant and undoubtedly data.

The aim is always the same, to provide the business the trustworthy insights in order to take the right decisions in terms of products, customers segmentation and future investments according to the strategy established by the company.

2.2 Data Collection

The concept of data- driven involves also all the methods by which the company is able to collect the data, to process it. Now days the human has realized the importance that data has in the business world, we have arrived to the point where it is possible to do data trade because throughout a data that is timely, accurate and accessible is possible to measure the actual performance of a company, to obtain some inferences about the future market movements and to model the future behavior of the customers.

Is incredible the power that technology has acquired, being possible to model the customer behaviors, because through an study of the data available and the acquisition of context analysis is possible to identify a given pattern, the idea is not to remain stuck in the data collection process but to allow this information to be accessible to the different competence areas of a company, when a data is coherent with the core business, complete and consistent is necessary to make it available according to each business unit, so moreover clean than and accurate also the data must be:

- 1. Joinable (should have a logic already defined where the data and metrics are correlated and also there must be a model already established where the vlookups problem of Excel won't be evident.
- 2. Shareable, This is a matter of culture within an organization.
- 3. Querable: the company should be able to "slice and dice" the data in order to convert it in a valuable resource in the decision making processes. We will talk about all the tools that are disposable in order to derogate this kind of activities also talking about *SQL language* and data base.

It is important to remark that obtaining the data and ensuring the veracity of it, is really complicated, the industry and all the high level standards apply to it are evolving and in order to survive among all the competitors is a key factor to follow the market requirements, but now days the company's are carrying so much about the value chain, all the productivity activities and the experience offered to the clients , but what happen whit the information left to effectuate all the corresponding analysis of the facts? are they ensuring the unbiased of the information or the timely of it? It's obvious that at this point the process of massaging the data has been able but what about the cleaning process? all the information that has been gathering is useful for the sales and marketing area? or it is taking into account information that make people to take the wrong decisions? the data is exposed to human mistakes because normally in almost all the processes in a company there are too many activities which involves human interaction, when they have to Transcript data from a given event trough an application there can be different types of errors such as *Insertion, Deletion and Transposition*

This is a business process argument which involves data quality methods that have to be applied and other kind of techniques such as elimination as possible of the human interaction, reduction of the steps and creation of different validation fields where is possible to elaborate a check and ensure in a high percent the data reliability. Everything oriented to deliver the information related to the right time , place and form (structure). reducing at the maximum level the dirty data and the anomalies.

The quality of the data is not just a concern of the Information systems department is a concern of the whole organization members, is a matter of responsibility because a wrong management of this data will lead to a significant loss in the company profitability account, because of that data quality shouldn't be left just to the analyst, data engineers, or CEO'S.

"Managing the decision-making process in a company is a crucial part of maintaining a wellfunctioning organization which is why much more attention needs to be directed at how decisions are being made"[23] how organizations can make better decisions and execute on them in a clearer, more efficient way. There are an innumerate possible decisions that can be taken thanks to the data collected during the sales and operation processes *Core Business*, some examples to illustrate the argument are evidenced:

- The Marketing department base it's sales campaign following the trends that are evidenced during the last month consumers behavior.
- The clusterization of a certain group of clients is made under the customer preferences and allows to take decisions like what kind of discounts offer to them based under their behavior, and what kind of customer card to offer them.
- The purchasing department localized it's strategy based on the customer behavior, the period of the year the pass sells data.
- According to the sells strategy of a productive company if is Pull or put centralized demand is possible to make some forecast about the number of units to be produced during a certain amount of time, also the number of resources to be utilize and the people needed in order to arrive to the different production objectives of the company.

What should be inferred from here is that company's should collect data answering to the right questions in order to identify which in specific will be the data that should be classified as core. Normally the current companies should think in terms of collection of data under the tree V's paradigm

- Volume: This dimension refers to the amount of data that a company should manage, and in this case is really important to understand the company infrastructure in terms of machines and servers available in order to gather and store data.
- Variety: in this case we also talk in terms of cost, because the more varied the data source and data types the costly that will be the infrastructure.
- Velocity: refers to the amount of data that must be processed per unit time.large-scale and real time processing is complex and costly.

This is a matter of trading -off, the company should take into account it's strategy and make available all the resources in order to accomplish the target, this includes get deeply in terms of one or other dimension, but always considering the opportunity cost, this is a clear way to prioritize the data sources for consumption and provisioning.

As has been defined early the company's data represents a really valuable resource for the own company but also for others such the competitors or other companies that has complementary products on their core business, and moreover than buying data, the company could opt for the managing data outsourcing, this is an strategy non recommended because of the power of data but in the case the company because of experience and different interest decides to do it, the important takeaway from this section is to buy or gather the raw data , not aggregated because every day there are future opportunities that could be analyzed through data collected previously that in the past where not considered as essential and in the future could represent the definition of new market chances.

Thousands of examples about the manage of data to identify the context of the customers has been seeing in the literature but there is a particular example done by "Creating a Data-Driven Organization" from Data Day Seattle 2015 as the Figure represents and is evident how a simple research online can derive into a deeply understanding of the customer preferences , in this case shows how a girl (customer) buys a product on amazon an through a simple login is possible to identify friends, address and house needs, opening new markets to the information holder.



FIGURE 2.1. Example. Importance of context. Figure reproduced from [6]

Apart from all the characteristics that have been determined during this table of concepts it is relevant to talk about the data governance and how all the records should be done in a standardize way and documented in order to create an unique language around the company to the different areas of competence, this can be done throughout a data dictionary that can take format with help of some information tools on the figure (2.2) is possible to see the data dictionary

പ്രം	Ro - Actions						
90	COLUMN NAME	DATA TYPE	NULLABLE	DATA DEFAULT	COLUMN ID	COMMENT	
1	FORN_ID	NUMBER	No	(null)	1	(null)	
2	FORN_CD	VARCHAR2 (12 BYTE)	No	(null)	2	(null)	
3	FORN_DESC	VARCHAR2 (100 B	No	(null)	3	(null)	
4	FORN_PIVA_CD	VARCHAR2 (50 BYTE)	No	(null)	4	(null)	
5	FORN_DATA_INI_ID	NUMBER	No	(null)	5	(null)	
6	CARREFOUR_CD	VARCHAR2 (13 BYTE)	No	(null)	6	(null)	
7	DATA_REPLICA_ID	NUMBER(13,0)	Yes	(null)	7	(null)	
8	INSERIM_DATA_ID	NUMBER	Yes	(null)	8	(null)	
9	CONTR_ID	NUMBER	No	(null)	9	(null)	
10	AGGIORN_DATA_ID	NUMBER	Yes	(null)	10	(null)	
11	FOAM_CD	VARCHAR2 (8 BYTE)	Yes	(null)	11	(null)	
12	FOAM_DESC	VARCHAR2 (60 BYTE)	Yes	(null)	12	(null)	
13	COND_PAG_CD	VARCHAR2 (4 BYTE)	Yes	(null)	13	(null)	
14	PRES_LISTINO_FLG	CHAR (1 BYTE)	Yes	(null)	14	(null)	
15	BLOCCO_CD	VARCHAR2 (1 BYTE)	Yes	(null)	15	(null)	
16	BLOCCO_DESC	VARCHAR2 (50 BYTE)	Yes	(null)	16	(null)	
17	CONTR_CD	VARCHAR2 (12 BYTE)	Yes	(null)	17	(null)	
18	DATA THE TO	NUMBER	Vee	(2011)	10	(20111)	

of the Iper Montebello providers as an example of trace-ability.

FIGURE 2.2. Iper Montebello Data Dictionary. Figure reproduced from [15]

From now on and for the purposes of this dissertation will be followed the model of Dykes that can be appreciate on the figure (2.3) in which is possible to identify what he considers as the analytics value chain, which starts on the data collection and finish with the transformation of this data in order to act as a critical evidence to help inform and influence strategy.

So far, has been described the Data collection importance within the companies, it is expected to bring into the table the concept of *Reporting*, which is the base of this dissertation and also the base of the rest of the concepts here described. Further on, the concept of applying Analysis processes within the company (2.4), Then, the theory of Decision Making process (2.6)



FIGURE 2.3. The Analytics value Chain. Figure reproduced from [13]

2.3 Reporting

Business reports, documents the progress of the organization, is a tool that allows the user to compare according to the metrics established the performance of a certain area or resource within a company among different time periods, the data that is collected in reports can serve a number of important purposes. As business grows there are countless reasons in order to effectuate a report and to record it historically in the company documents as a paper trail of the past.

The reporting is an activity very used inside the companies in order to illustrate the facts, but is important to highlight that in the majority of the cases, is about a number without value, in order to provide value to this type of numbers there is necessary to find the context in which this information has been collected with the aim of giving it coherency according to the business, the target is to go forward from the no causal explanation above the descriptive analysis that will be describe deeply in the next section (2.4). Now is described the argument about how to prepare the information to the final users making reference on what is happening right now?.

Reporting is a fundamental part of the larger movement towards improved business intelligence and knowledge management. Often implementation involves extract, transform, and load (ETL) procedures in coordination with a data warehouse and then using one or more reporting tools. Reports can be distributed in print form, via email or accessed via a corporate intranet. [14].

2.3.1 Metric Design

Metrics are used in order to identify what is important to measure so it is possible to evidence whether the organization is achieving the goals previously established and to drive all the strategies in order to improve the indicators. Helping the business to focus on what is really important, a common mistake done by the companies is to formalized to many objectives that sometimes are really difficult to achieve all at the same time, what it is recommended is to identify the key goals at the short and long term but that could be reached by the people and resources with a more realistic approach. Overall, Metrics should reflect all the business requirements including also all the information required by the law for legal, safety and contractual purposes. Metrics are the mirror of the company priorities.

According to the six sigma methodology the definition of the metrics within an organization should be clear so it could be possible to bench market the success of the company and to setting out the right strategy determining a set of metrics in order to track whether the business is heading in the right path and to monitor success. The best practices established that so as to define the right metrics it is important to make them simple so it is possible to *transfer* the knowledge and the procedure defined in order to measure it, it is really simple to *implement* and goes according to the savvy business mentality and last but not least to allow the metrics to be *comparable* within the company, but also to make clear statements about the situation of the company respect to it's competitors.

So at the end, the target is to reduce confusion, making an standardization throughout the collaboration of the different teams inside the business areas making it centralized, automated and documented. Building up a repository of the institutional analysis about the causal and factual factors in the business.

It is possible to evidence on the fig (2.4) as an example of metrics under the marketing context, how metrics should be done below a certain logic they should be *accurate* in order to

take in consideration all the possible bias that could be presented during the measurement process and moreover the data it self how can suffer the human countless way to introduce errors, furthermore should be *precise* here it is important to analyze the size of the sample because as statistics establish the larger the sample the smaller the standard error which means that the data measured in a a different time-line period should return more or less the same information. *Robust* and *Direct* are other two qualities that metrics must posses because in this way they are insensitive to extreme values outside of the quality control chart, besides it also is necessary to ensure the measuring process in order to verify that the information collected describes the underlying process

Define Objectives and Metrics Here is an example for business stages mapped to the buyer's journey							
Phase	Objective	Strategy	Program	Metrics	Business Impact		
Awareness	Drive awareness, early interest	Expose target audience to select content	Economist content program (mix of PR, social, SEO)	Impressions, reach, increased views on site	Ex: SOV ,ToMA		
Engage	Generate initial demand	Multichannel campaign focusing on key findings of Economist	PPC, email, social, luncheon with Economist speaker	Downloads, new names, CPL	Engagement with select companies and titles		
Convert	Drive conversion, deal acceleration	Focus on how MA helps address key findings	Nurture program Late stage accelerators	#MQLs, #SQLs #opportunities Pipeline	New pipeline created, pipe coverage		
Retain	Delight customers	Focus on helping customers understand strategy + how to tips	(webinar + email) – showcase series	Usage data, # partner integrations	Sentiment and satisfaction, >90 % retention, lifetime value		
Advocacy	Inspire evangelism	Identify happy customers showcase their results	Purple Select	Likes, comments, retweets, satisfaction	Referrals, reviews, positive word of mouth, NPS		

FIGURE 2.4. Illustration of the metrics within a company. Figure reproduced from [6]

Moreover, than just defining the metrics in order to deliver real value and progress it is necessary to invest on the methodology for data collection that also needs to be standardize across the different departments and functional areas of the organization, in order to finally arrive to a process of sharing the data where the takeaways are expected to be continual part of the improvement. that means that as the customer requirement changes around time and the offer grows exponentially also the internal processes and "modus operandi" will change, so it is necessary to implement within each organization a process in which the metrics that were already created should be re-evaluated according to the current strategy and procedures.

At the end, under a given structure, metrics can be a tool in order to challenge the company, some enterprises in the market have been established some incentives to their employees in order to arrive to certain targets this motivates people that are goal-oriented.

One way to keep metrics understandable is to use the SMART model for institute KPI'S (specific, measurable, achievable, relevant, time-based) model. All the attributes have been

explained during this dissertation chapter but in this particular discussion is relevant to stabilize a target really clear an precise so it is possible identify Why?, How?, When?. so the stuff can have access to a given feedback.

Financial performance	Understanding customers
 Net profit Net profit margin Gross profit margin Operating profit margin EBITDA Revenue growth rate Total shareholder return (TSR) Economic value added (EVA) Return on investment (ROI) Return on capital employed (ROCE) Return on equity (ROE) Debt-to-equity (D/E) ratio Cash conversion cycle (CCC) Working capital ratio Operating expense ratio (OER) CAPEX to sales ratio Price earnings ratio (P/E ratio) 	 Net Promoter Score (NPS) Customer retention rate Customer satisfaction index Customer profitability score Customer lifetime value (CLV) Customer turnover rate Customer engagement Customer complaints

FIGURE 2.5. Subset of common KPI's. Figure reproduced from [4]

As the Figure (2.5) entails the most common KPI's measured in the industry are almost standardize and also some of them are required by the law, alongside it is important that each company creates tailored a set of KPI's based on the core business, targets and the strategy, using a sensory and specific language.

2.4 Analysis

Retaking the argument of the Analysis value chain from Dykes (2.3) so fare we have localized into the Descriptive situation analysis of a company throughout different tools as collection of data section(2.2) definition of metrics subsection (2.3.1) to extract an track the valuable information related to each process within the organization and due to those KPI's established is possible to recreate reports facilitating the view of the information, all this tools allow the user to identify **What happened?** studies have demonstrated that the majority of the incumbents present on the market has access to this information what changes from now on will correspond to a significant acquisition of competitive advantage, referring to **Analytics** inside an enterprise. The Process of Analytics will provide the company of a completely new vision where the CIO's will have access to a prescriptive analysis where the solution to the next question could be find**Why it happened**? it will concede the construction of the context, the storytelling that will be explained deeply in the next chapter, privileging to dig down the information. The final Target is to influence the decisions - making process and to maintain informed all the competences areas, to create an impact understanding specific cause- and - effect in a certain context.

but who is the people prepared to effectuate this kind of job?, which type of characteristics and background studies are necessary in order to structure a mentality making it able to ask the right questions? according to the 6 sigma method there are five questions that needs to be done in order to achieve the target on each specific occupation.

"5 Whys is an iterative interrogative technique used to explore the cause-and-effect relationships underlying a particular problem.[1] The primary goal of the technique is to determine the root cause of a defect or problem by repeating the question "Why?" Each answer forms the basis of the next question. The "5" in the name derives from an anecdotal observation on the number of iterations needed to resolve the problem.." [10].

2.4.1 Type of analysis

Like this methodology there are many more which allows the user to gain some insights about the raw data collected. According with [20] there are six type of analysis that can be done which are described as follows:

- **Descriptive**: This is the simplest type of analysis, describing a univariate data, describing a single data; there are a lot of type and common measures that enable the user to display data on the dashboards.
 - 1. *Sample Size*: A sample size is a part of the population chosen for a given analysis .and describes the number of data it is important to defined wisely the sample size in order to have direct data that makes reference to an underlying process. Finding a sample size is one of the tough activities and depends upon many factors.
 - 2. *Mean*:It is an average where there must be add up all the numbers and then divide by the number of numbers.
 - 3. Median: The median is the middle number.
 - 4. *Mode*: number that is repeated the most.
 - 5. Minimum: The smallest Value in the sample.
 - 6. Maximum: The greater value in the sample.
 - 7. Quartile: The central of the data.
 - 8. Range:Maximum value on the sample minus the minimum value on the sample.
 - 9. *Standard Error*:Measures the expected standard deviation of a sample mean if were taken repetitively measurements with the same sample size of the original.

10. *Variance*:Is a measure of dispersion, and a lot of inferences can be taken from this metric, generally identify whether a sample is trustworthy.

The key of this type of analysis is that the interpreter should be able to slice and dice the information describing the key features of the sample numerically. and is normally consider the basis for the next type of analysis.

- **Exploratory**: Data Analysis (EDA). At this point of the data analysis is important to reflect which kind of data the user ha available and then to pose the right questions is possible to do it by looking at patterns, trends, outlier. In this way is possible to get to the context, to find a rational sense to the story, so you can get some clues about which are the next steps in order to take advantage of the timely and accessible data, according to [30] the following are a cast of activities that can be done due to the *EDA*:
 - 1. Maximize insight into a data set.
 - 2. Uncover underlying structure.
 - 3. Extract important variables.
 - 4. Detect outliers and anomalies.
 - 5. Test underlying assumptions.
 - 6. Develop parsimonious models.
 - 7. Determine optimal factor settings.

Most EDA techniques are graphical thanks to the open-mindedly explore characteristic of the them, which gives the analysts a magnificent power to go beyond just a number.

- **Inferential**: the predictions are based under the inferential analysis. Through past events or given samples normally the organizations make some generalizations about a population. There are two main areas of inferential statistics:
 - 1. Estimating parameters: This means taking a statistic from your sample data and using it to establish a population parameter , this is the where descriptive, EDA ad Inferential analysis collides.
 - 2. Hypothesis tests the idea is to infer some information as parameters, distributions or relationships .Inferential statistics use statistical models to help you compare your sample data to other samples or to previous research thanks to statistical models called Generalized Linear model As an example of this models we can take the Anova table or the t- Student test which also allows to make a deep reasoning about the confidence that can be inferred from a particular sample , also is possible to reconstruct the model through a linear model and identify which are the variables that influences an underlying process The illustration (2.6) refers a study made by [8]

where were analyzed some variables that could influence the politics knowledge. In this case they analyzed if the gender and even the education level are correlated to the level of politics knowledge.

Dependent Variable:Int_Politics							
Gender	Edu_Level	Mean	Std. Deviation	N			
Male	School	38.2000	4.18463	10			
	College	44.1000	4.26745	10			
	University	64.1000	3.07137	10			
	Total	48.8000	11.87841	30			
Female	School	39.6000	3.27278	10			
	College	44.6000	3.27278	10			
	University	58.0000	6.46357	10			
	Total	47.4000	9.05767	30			
Total	School	38.9000	3.72615	20			
	College	44.3500	3.71023	20			
	University	61.0500	5.83524	20			
	Total	48.1000	10.49649	60			

Descriptive Statistics

FIGURE 2.6. Generalized linear model, Anova. Figure reproduced from [8]

Inferential analysis are made because allows to predict a certain behavior but is also less expensive than sometimes even the data collection, inferential analysis is when you try to infer from a sample that you have to a larger population.

• **Predictive**: the predictions are based under the inferential analysis. Through past events or given samples. Normally, the organizations make some generalizations about a population but beyond that, predictive analysis is used in order to generate forecast, future prediction in a time series. Currently on the market this kind of analysis are used in order to forecast the demand, hence the production units per a certain period, therefore the resources needed and the capacity of the manufacturing process. One example of predictive analysis is the theory of tails which allows to simulate the number of clients and the service capacity mean a given service is delivered. an example can be found on the



FIGURE 2.7. Flexsim software simulation process. Figure reproduced from [9]

• **Causal**: After doing all the previous analysis we are able to find a really high correlation between variables but unfortunately this doesn't mean the causality of the event, in order to reduce the uncertainty various experiments and test must be done changing in different ways the factors so becomes possible to control the result.

After the implementation of some of the previous analysis is possible to:

2.4.2 Levels of Analytics

There are further distinctions within each of the five categories of Analysis that are useful for applying models and using tools in order to manage the information. according with [17] it is possible to identify eight labels

- 1. Standard Reports
- 2. Ad hoc reports
- 3. Query drill down
- 4. Alerts
- 5. Statistical Analysis
- 6. Forecasting
- 7. Predictive modelling optimization

Each level of the data-driven approaches increases complexity and power over the previous one. as can been seen on the figure (2.8).



FIGURE 2.8. Analytics Levels. Figure reproduced from [17]

Improve the **mindset** and the optimization of the core processes trough the inference an predictive analysis is possible to make some forecast on SALES. Gives a higher power to the functional analyst in order to choose under a suite of weighted variables, so the resource can take decision about the information to show and from which to make the analysis.

Causal modeling draws from both data-driven and expert-driven techniques. On the next diagram it is possible to identify how the different type of analytics are used in order to make different levels of analysis, is the joined reasoning of the arguments treated during the subsection (2.4.1) and the subsection (2.4.2).For instance, the inferential type of analysis can be used to prepare a forecasting or better it can be use to derive an optimization of a given process.



FIGURE 2.9. Intersection between the level of analytics and type of analysis. Figure reproduced from [17]

2.4.3 Analyst human component

The key success factor within a company are the competences that have been brought by the human resources acquired previously. In order to make an adequate analysis of the data and to generate a real change on the management and procedures within a company, It is necessary to create conscious about the business requirements and the new technologies and techniques available in the current market.

"Machines don't make the essential and important connections among data and they don,Aôt create information. Humans do. Tools have the power to make work easier and solve problems. A tool is an enabler, facilitator, accelerator and magnifier of human capability, not its replacement or surrogate. A fool with a tool is still a fool." [29].

With the stimulation of the millions of tools available on the market and the learning machine age, the companies are investing all the resources in order to obtain the latest technologies, but they are not investing to form and train people, they acquire new graduated employees, with the methodology defined but without the savvy business competence because they don't have experience. The right approach is to create an equilibrium.

human component is often left aside due to the excitement over data tools. Consider how we talk about Big Data. We forget that it is not about the data; it is about customers having a deep, engaging, connection with the products and services offered, so it's a matter of expertise and trajectory to eliminate as much as possible the risks.

2.4.3.1 Roles

Furthermore, a company requires to have inside different teams composed by a certain roles definitive when is time to talk about analysis.

- **Data Analyst**: This role is the basic one which is in charge of interpreting data and converting it into valuable information using all the analysis techniques reviewed during the subsection (2.4), they have also responsibilities related to the quality data, ensuring about the data coherency and accuracy, they should facilitate and implement a methodology in order to optimize the data collection processes.
- **Data Engineers**: They are basically in charge of render the information available and cleaned to the analyst, so the can make the respective analysis, they can also be responsible for the business intelligence tools that are provided as enterprise resources, so they can
- **Business Analyst** : This is a really important figure inside a company because it works as bridge between the it department and the other areas, has to act as a filter about all the requirements of the specific sector helping to defined the metrics, understanding the KPI's and the target so each process is documented and there is a logic defined, the idea is to synchronize this figure with the other analyst internally so to make sure they are talking the same business language.
- **Data Scientist** is the closer role to the technicality, should be really good managing software engineering and statistics.
- **Statisticians**: Resource Modelling skilled focus on the metrics and with the use of the company tools make inferences in order to help during the decision-making process.
- **Quants**: This role is really appreciated among the financial sector in order to model securities, trade assets, and price options and have to be mathematician skilled.
- **Data Visualization**: If the company also wants to be specialize in terms of visualization aesthetics creating info-graphs and dashboards should acquire the competences of a person with a previous background doing this kind of job.

Obviously, the people that are needed inside the company depends on the business strategy and the objectives established also the market segment in which the company is classified. In many businesses, it,Äôs difficult to define individual roles and responsibilities. Unfortunately, this can lead to a great deal of confusion within the company. Clearly defining roles and responsibilities will have a positive impact and it will be reflected over the results that different projects could bring, as all the previous defined roles have some activities in common or that could be sharing , the organization should prioritize.

2.4.3.2 Skills

By way of summary, all the respective roles that are needed within a company should own different type of competences and capabilities, they must be able to gather data, select them, design right metrics, create processes and make testable predictions, but overall, they should have the right skills in order to ask the right questions. They should be data oriented since obviously they are expected to change a paradigm inside the company, the need is represented by a appropriate methodology, analyst are people skeptical confident because of their curiosity they should gain insights but after understanding the cause and effect they should be able to convey this information to the rest of the company, selling insights is not an easy task so one of the main characteristics is to be good communicators and storytellers, as the argument is to break the chasm and is not really easy considering an traditionalist enterprise and in which the employees have a certain jealousy about their knowledge and the idea of transferring them.

So, an analyst due to the human interaction that is fundamental in the process of data collection has to be patient, data lover, pragmatic and life learned because this person should be able to adapt in different contexts and work with different people and tools, they should be the enough determinant in order to kill a story that should end and implement it if it is the case because after selling the idea they should have the capacity of construct the solution and integrate it with the business in order to create and impact and influence in the determination process

2.5 Analytics Organization Structure

Continuing with the analytics value chain described in the figure (2.3), each company has to act, has to internalized all the competences that has acquired, even has to take advantage of the abilities within each area of the company. An Enterprise is organized according with the roles, skills and team structures, this is a managerial decision that will be determinant on the performance of the projects developed and could be the reason of numberless failures or profit-abilities so, how the analytic team is structured is a fundamental decision that depends on multiple variables always corresponding to a given context. The Incumbents are organized currently making use of two main data team structures, *Centralize analytics* (2.5.1) data and talent residing in one group or *Decentralized Analytics* (2.5.2) where it is share among the different functional areas. Not always should be analyzed the extremes, this is the case where there is a flexible option too, the *hybrid model structure* that will be see more in detail during the section (2.5.3). This types of organizations can be seen in all the types of structures, to illustrate the figure there is a case study on the health industry market



FIGURE 2.10. Health, Case study. Figure reproduced from [5]

2.5.1 Centralized

The Centralized approach stands for a team which is in charge of standardizing procedures, information , metrics and the skills training methodology along with the tooling. It's an economy of scale where a given group share the resources usable, allowing the reduction of process cost, training time and software licenses. A unique team must switch among all the different areas and range business. Talking in terms *Objectivity*, this concept is maximize because as external team a centralized structure doesn't have any particular interest on other projects inside the company, so they will try to reduce the customization as much as possible, to promote master data as a single source of trust.

"Everything needs to get into the pipeline and get prioritized, and get resources allocated against it" [19].

As Piyanka [19] in his articles remarks that the structure of analytics in large organizations can take many forms, but the centralized one at the same level that can contribute to an improvement inside the organization also has some flaws, from the fact that the generalized situation includes human resources that rationalize following a technical approach and often are removed from the business owners from their goals and are not taken in to consideration when taking decisions. They are set in order to configure the whole organization analytic methodology and this without the help of a business entity could be a really big problem, depending also on the curiosity level of their participants in order to understand the company but it also means to leave aside some arguments that in a given point could be important and a reduction on the flexibility level.

As reported by the figure (2.10) the centralized model is prone to suffer from lack of information, there could be a lot of miss understandings when not all the other areas are aligned with the centralized analytic team.

2.5.2 Decentralized

The decentralized approach, each business unit is able to recreate a methodology of analysis tailored in agreement with the business needs so it will agile for the immediate information, this is an easy way when we talk about internal management of the information, there won't be a gatekeeper neither a filter of the information, so the execution depends exclusively on the person in charge, this allows the team to make test and experiment with the data , this will lead to a better understanding of their business areas and a preparation in order to ask properly questions. they will identify the core data but always independent of the whole business organization, representing also a increase in the maturity process under a roof of the company.

The people that are directly connected with the business, they are specialist who knows and understand the goals, the metrics and reports needed, but maybe they don't have the competences to establish which are the best practices in order to use a tool, and even worst the wont be coordinated with the whole business. This lack of alignment will drive to a redundancy of the work and not right utilization of the resources, introducing a bias to make the data or work look better according to the interest of each project.

2.5.3 Hybrid Model

Whenever an organization starts to be aware about the importance of managing correctly the data and starts to become **data-driven** to achieve higher levels of efficiency and competitive advantage. Using a hybrid model that stands the unification of both approaches *centralized* described on subsection (2.5.1) and *Decentralized* (2.5.2) taking the objectivity and the flexibility concept where there will be a filter applied on the business demand. bringing analyst together to the development process combining the clear career path of a centralized model with the faster turnaround time and the direct, full time access to the information of the decentralized model allowing to gain to experiment with new types of models, to test, customize .A hybrid analytics strategy will ensure that you have the right tools for any job and a data governance and promotion.

As the research made by [5] exhibit on the figure (2.10) about the health care case study, Both entities work together to ensure that the HIE can track a patient,Äôs movements between participating providers while ensuring that health-care organizations can receive a minimum clinical data set for each beneficiary. This is a model that in the case of the health care industry has produced long-term reward for many.

2.6 Decision Making

Supposing that the company posses unbiased , relevant, timely and accessible data managed by very skilled and specialized people where the data is aligned among all the organization and the tools needed in order to visualize the information are made available. Studies conducted have demonstrated that the fundamental driver in order to take decisions of the CEO's is the own intuition and experience, they could be a informed about the data and current results internally in the company also they can be a bit influenced by data but normally the decisions are subjective according what they have seen among the expertise. the data sometimes can tell about the direction that a company should take but in some cases and depending on the context the company cant follow that path because of the strategy, goals and morality of the organization so there should be a mix. The data in this case will generate questions and deliberations rather than point to a specific decision option. The takeaway here is to be data based analyzer and focused.

Chapter 3

Theoretical Framework

3.1 System Architecture

To create a successful system of systems it is necessary to meet the needs of it's stakeholders and to evolve and scale to sustain all the data process implemented at the inside of a company, rendering possible to increment the flexibility and the optimization of the process challenge. In a enterprise where there are different type of systems that should communicated and interact, the biggest challenge is to achieve a cooperation between them ,mean, each of them works as an entity, normally in order to reach this criteria is necessary to make some integration activities that could guarantied the correct data transmission as will be seen deeply in the subsection (3.1.1)

There are too many problems when it is about integration also, when there is an specific architecture already implemented and what is intended to do, is to integrate a new system into the system of systems, in order to identify the integration pattern is important to understand the perimeter of each integration intended as how data will be shared and how control will be managed. [28].

There exists different type of Architecture inside a company and even if there can be combination of the main ones for the purpose of this dissertation we will describe the two most common:

• *Spaghetti Architecture*: This type of integration involves the countless number of connections and the generation of a work logic that will be really difficult to manage every time that grows, degenerating into an exponential tangle of connections and involves increasing complexity and costs of maintenance and evolution (spaghetti integration). With the growth of business applications and the need to connect multiple applications in the development of Web applications, in the field of integration tools it is necessary to generate a change in terms of cost and fluidity: that is why, compared to "tactical" approaches to The EAI (Enterprise Application Integration) platforms were created specifically for the specific



point-to-point needs, distinguished by the presence of an integration Middleware that allows a centralized, disciplined and more standardized integration. [21]

FIGURE 3.1. Spaghetti Structure. Figure reproduced from [21]

This time of architecture entails a complicated collaboration between the systems within a company, there is a Register for each entity and the level of efficiency is really reduced

• Organized Architecture:



FIGURE 3.2. Organized Structure. Figure reproduced from [15]

This type of integration between systems is made, trying to generate an optimized sorting

of the information, increasing the efficiency mean the convey the data among the systems through the use of a Middleware. Who acts as an unique central role becoming possible the normalization process among all the possible systems. A service-oriented, flexible and decoupled IT architecture reduces project maintenance time and costs, as well as the need for customization and evolution of the same against new integration requirements.*SoS*

3.1.1 Data Transmission

There are different structures by which it is possible to allow a correct data transmission to build a customized and automated integration that allows to administrate, trust and secure file transfers.

MQ is an IBM solution where there are two agents, each connected to their own agent queue manager, A file is transferred from the agent on the one side , to the agent on the other side. Applications and tools connect to these queue managers to configure, administer, operate, and log IBM MQ Managed File Transfer activity in the IBM MQ network.[8]

FTP is a File Transfer Protocol in which the client gets the data relying on two communications channels between client and server: a command channel for controlling the conversation and a data channel for transmitting file content. at the end the final user can insert, download, eliminate, change the name, move and copy files on a server.

Web Services is present in the internet and makes available information trough a XML. *extensible Markup Language* metalanguage is a markup language that makes use of a static mechanism that allows to the meaning of the elements inside a document. So the *XML* servers to encode the communications addressing it to a web service. Those web services are not Colgate to any type of language being self-contained, that can be invoked over the network.

3.2 Analytics Maturity

The maturity of an organization referred to the concept of analytics is given by the level of expertise and domain about the concepts, knowledge, best practices, methodology, and implementation process integrated within a company in order to derogate better services increase the profits and achieve all the targets established.

The mature is measure thanks to the stability applying processes that exits or are used inside a company normally, the level of analytics maturity is also highly correlated to the global organization maturity, assessing the people, the process, the technology and the Measurement intended by the tools used to collect data and to interpret it. In the table bellow is possible to identify the concepts treated during the chapter (2) about the level analytics and the type of analytics that are applying within a company, those are concepts that defined the level.

	Level 1	Level 2	Level 3	Level 4	Level 5
	Performed	Managed	Established	Predictable	Optimizing
People	Success depends on individual heroics. "Fire fighting is a way of life." Relationshi ps between disciplines are uncoordinated , perhaps even adversarial.	Success depends on individuals and management system supports. Commitmen ts are understood and managed. People are trained.	Project groups work together, perhaps as an integrated product team. Training is planned and provided according to roles.	A strong sense of teamwork exists within each project.	A strong sense of teamwork exists across the organization. Everyone is involved in process improvement.
Process	Few stable processes exist or are used.	Documente d and stable estimating, planning, and commitment processes are at the project level.	Integrated management and engineering processes are used across the organization.	Processes are quantitatively understood and stabilized.	Processes are continuously and systematically improved.
Technology	The introduction of new technology is risky.	Technology supports established, stable activities.	New technologies are evaluated on a qualitative basis.	New technologies are evaluated on a quantitative basis.	New technologies are proactively pursued and deployed.
Measurement	Data collection and analysis are ad hoc.	Planning and management data is used by individual projects.	Data is collected and used in all defined processes. Data is systematically shared across projects.	Data definition and collection are standardized across the organization Data is used to understand the process qualitatively and stabilize it.	Data is used to evaluate and select process improvements.

FIGURE 3.3. Organization Maturity. Figure reproduced from [3]

The importance of taking decisions based on trustworthy data is now a necessity but despite this, isn't clear how the companies are applying it and the steps that they should pass in order to be classified as experts, those steps are classified in different stages that conforming to [17]
accessible, high-quality data enterprise orientation analytic leadership strategic, targets and analytics.

Data quality is one of the main characteristics right now with a higher level of importance because of the interest from many companies about collecting as much data as it is possible, so this factor becomes decisive in order to identify maturity levels.

There are many variables that could influence the analytics maturity of an organization, the most important thing is starting to diffuse this concept and understanding which are the attributes that allows to go through all the stages. Cause in this way it will help the organization during the process of evaluating their own competences and the way they are used in order to achieve a coordination between the It areas inside a company and all the other departments in charge, increasing the profits and improving the customer services.

The majority of enterprises, since they consider the technological maturity fundamental in order to complete all the analysis needed for a good performance are trying to search about a model that en gloves and gives a direction to follow. Practically, there are different approaches took about The analytics maturity some of them are:

3.2.1 Technical Approach

The technological approach usually includes the Data Warehousing Maturity Model developed by [33]

A **data warehouse** is a specially prepared repository of data designed to support decision making. there are different sources of data and file transfer protocols *FTP* internal systems could be also feed by other external sources this is the case of companies that externalize some process and the want to internalize the data collected. To create the data ware-house *DWH*, is necessary to draw the data from diverse operational systems. Each Data Ware House has his own data Base *DB* in order to provide a set of data that supports decision making (chapter 2). Once the *DWH* is created and consolidated people inside the company with the purpose of analyze the information and with a given formation like Data Analyst, Business Analyst, Data Engineers or Data science, they should have access to the data trough a *SQL Language* 3.2.1.1 because is made under the concepts of relational data which provides a higher level of efficiency. Based on the stages of growth theory, the Data Warehousing Maturity Model includes three stages:

1. *Initiation*: Stage intended to describe the first version of the *DWH*, it is important to stand out the fact that implementing a DWH structure initially provides a significant reduction on the IT cost and personnel cost, because also leads to a dismiss of the searching and analysis processes, this stage is when the company starts to realize the flaws and a wide window full of opportunities starts to open, here is when the vision of the enterprise starts to amplifies.

- 2. Growth When a company arrives til this point means that there was a POC Proof of concepts, that only could represents a procedural improvement, a reduction on time, a cost optimization, After the process of identifying the advantages the company could suffer many risks, the principal identified is known as $, \ddot{A} úsilos of information, \ddot{A} \dot{u}$ this is a problem that could be presented between the different teams where there is fear about to share the information, so there is not a fluid of information among the company instead there is a impossibility from the system to work unrelated systems, this phenomenon occurs specially when each team works with an specific set of data, here is evident the redundancy that could block the vision of the company in the search of new opportunities.
- 3. *Maturity*: In the maturity stage, the volume of the data maintained grows, it covers multiple subject areas, it is highly detailed, and it provides considerable historical detail, referring to [3]

The Maturity of a *DWH* is defined by nine mostly technical characteristics of enterprise data warehouses that are exposed in the figure 3.4:



FIGURE 3.4. Maturity Stages of DWH Figure reproduced from [3]

Here are listed the Top 4 Data Warehouse Tools and Testing Techniques:

- Amazon Redshift
- Teradata
- Oracle
- Informatica

3.2.1.1 SQL language

SQL stands for Structured query language, is a domain specific language used in programming and designed for managing data, based in a RDBMS relational database management system, this is a structure extremely useful when it is necessary to manage enormous amount of data with an unique command at the same time, specially when there are connections between entities or variables of the data.

SQL provides to the user many possibilities when it is about administrate data such as Data manipulation; it is possible to insert, update or eliminate data from the Relational Data Base, alL off this is made according an structure that must be define trough the data definition and last but not least the system gives to the user a million of possibilities thanks to the accessibility, usually there are different types of sub-languages used in order to do possible all the previous functionalities:

- 1. *DQL* Data Query Language includes the commands made basically SELECT statements. SELECT statements let you query the database and convey the results throughout a array structure according how the data was map since the beginning.
- 2. *DDL* Data Definition Language: This is the language that allows to set a given scheme to the data that will be store in the *DB* includes the statement of create, drop or alter.
- 3. DML Data Manipulation Language: what is left in order to fill the tables created thanks to the DDL language is the DML language which allows to INSERT, UPDATE AND DELETE, aside from this in order to select the information that wants to be collected by the user the language communicates also with the DQL language in order to make the SELECT statement.

SQL became a standard of the American National Standards Institute (ANSI) in 1986, and of the International Organization for Standardization (ISO) in 1987.[18] Since there is an standard must of the data bases can change a bit with the commands to manipulate the information. tables and query results are lists of rows where are displayed the information.

The *RDBMS* is also based on the Relational Model *RM* following a concept called first-order predicate logic, where all data is presented in tupples and grouped in relations, what is searched trough this model first planted by Codd [7] in 1969 is to provide relational operators to manipulate the data in tabular form.

The purpose of the relational model is to provide a declarative method for specifying data and queries:when users needs to access to the information they must state trough a command the place and the data that they want to visualized, and let the database management system software take care of describing data structures for storing the data and retrieval procedures for answering queries.[7]

CHAPTER 3. THEORETICAL FRAMEWORK



FIGURE 3.5. Diagram of an example database according to the relational model. Figure reproduced from [31]

The majority of the *DB* are based under the SQL data definition and query language so one as one of the main requirements saw during the chapter (2) was the querability of the data this one must be Joinable and Shareable, e in an SQL database there are presented different schema that contents key constraints, other constraints, and SQL queries correspond to predicates.

```
SELECT column1, column2, ...
FROM table_name
WHERE condition;
```

FIGURE 3.6. Example of SQL syntax. Figure reproduced from [18]

3.2.2 Non Technical Approach

Organization-focused analytics maturity models offer a non-technical view and discuss maturity from the business-technical perspective. To illustrate, Gartner,Äôs Maturity Model for Business Intelligence and Performance Management evaluates an organization,Äôs efforts in terms of dimensions such as business sponsorship, organizational structure support, scope of the analytics initiative and availability of performance metrics using a 5-level scale [27]



FIGURE 3.7. Gartner, Analytics Maturity. Figure reproduced from [27]

The idea of an organization is to arrive to the last level where it is obtained clean and timely data that influence business decisions throughout a complete internalization of the processes and with the help of specialized tools.

3.2.3 Information Systems

As was evidenced during this dissertation the IT area within an organization plays a radical role in order to propel the company maturity growth. Now days, they are integral part of the business decisions and influence the establishment of the strategy

IT's ability to successfully do so is dependent on how well the Business analytic alignment is achieved and the management commitment. As the majority of the organizations a centralized business analytics structure 2.5.1 is integrated inside the company ; but there is one thing clear and is the role that the managers plays regard to the encourage of the analytics processes.

As It decisions and resources are the company path determinants in the short term and long term, it is fundamental to chose the suitable systems that will concede to fulfill the business expectations this is also one of the variables influence in the score level of maturity inside the company as the performance and success as well.

How a company handles the information systems is the key for an organization to succeed, information systems not only make processing easier but they give value to any kind of company; held down the purposes of this dissertation were the focus is on analytics they are a main tool. The concept is defined in accordance with:

"Information Systems are interrelated components working together to collect, process, store, and disseminate information to support decision making, coordination, control, analysis, and visualization in an organization. ...In fact, we might say that one of the roles of information systems is to take data and turn it into information, and then transform that into organizational knowledge." [11].

When mentioning Information Systems, in the context of this section, it refers strictly to **Business Intelligence Systems: BIS** and *Management Information Systems: MIS*, these are the ones involving *hardware*, *software* and technology in general, with the aim of collecting, organizing, storing, manipulating, analyzing and communicating the information within an organization. The systems in mention boost the improvement managerial and the decision taking processes. thanks to the wide range of solutions and possibilities that disrupting the traditional system are proved to push the Enterprise efficiency and accuracy.

Because of the continuous evolving process of the information systems on the market, the continuous global research for new solutions to new requirements, some many changes have been applied to the processes the information systems should follow a given structure according to a process-oriented approach and is necessary so that the structure could also be adaptable to the changes following these characteristics: [32]

- *IS* should reflect clearly the structure of the business processes , in a easily way such that the user will reduce the number of possible errors that could be during the creation and configuration and during the normal service process.
- They should be able to integrate.
- The way in which are set up the Information systems must answer to the business changes so there will be certain level of flexibility that permits to modulate the system according to the processes
- They should allow the possibility of identifying diverse problems before the bug affects the company in a block way, also must allow changes and updates.
- The allocation of resources is also a really important point to make emphasis because more user friendly the platform is better the results. those are represented in the increase of performances and achievements of the KPI's.

The categories of the Information Systems are organized considering the functionality and the preparation of the resources in order to used them, *Office information systems*, *Transaction-processing systems*, *Knowledge-management systems:*, *Decision-support systems:*, *Control systems:* these different type of systems are classified considering what kind of activities can be taking under control doing use of them, the business knowledge that is gather and visualized by the

final user, the degree of specialization of the tool in a certain area, level of aggregation and the analytics processes that can be done with the use of the tool [32]

3.3 **Business Intelligence**

During this chapter there will be presented, more in depth some characteristics of the Business Intelligence and Data-Driven, giving also relevance to the Collection Data to be converted into Reporting and visualizations which are the center of the design of processes and on which is based this dissertation. First, will be describe the whole picture of Business intelligence intended as a methodology not just a set of tools in the section (3.3) getting into detail about its features and highlights that could be benefit for the companies and their importance and Moreover, in the Reporting and data discovery during the arguments on the subsection (3.3.4) it is presented the idea of how important it is to implement this framework to provide a common understanding between parts of an organization, presenting the main characteristics and specifications on the way these models are presented.

Companies now has seen BI as a useful tool that can help improving processes of analysis, right now it represents a crucial instrument to keep in track with what technological improvements that data management systems are bringing.

So far have been described the strategy, the people and knowledge that are needed in order to achieve the company strategy, but from now on and for the purposes of this dissertation, the emphasis will be among the concept of Business Intelligence and the tools that make possible to collect, process and analyze the data, allowing the company to obtain a context and to evaluate about the future decisions of the company. In the other hand BI will provide the tools to the customer in order to forecast the future, turning to be essential in order to increment the level efficiency, facilitating the communication process, the decisions process and so on.

Now days the abilities of the companies regarding to the processes of collecting and gathering data have been increasing exponentially but what to do with all of this information? how to identify the key components and behaviors of the business itself?, how to take the right data to make some analysis and identify some patterns in the society?, to achieve the most important goal of a company, to be an incumbent that creates competitive advantage and increase the value offered to the customer that at the end will be reflected in the increase of the sales and the profits for the stakeholders within the organization. **Business Intelligence** as has been defined by Dresner

Describe a set of concepts and methods to improve business decision making by using fact-base support systems [25].

Currently there are different tools to be used by the companies in order to communicate and allow resources inside the company to make all the analysis with a certain deepening level. Thus, it can be considered as an imperative framework, But is important to evidence that BI are not just tools, is a work method and structure that includes architecture, data bases *DB*, applications, Best practices *ITIL* and methodologies.

If the idea is to center the explanation in the specific advantages that provides the integration of a business intelligence process within a company, following there are enlisted some of the advantages to take into account when evaluating *BI*:

- Increase profitability: After doing more accurate analysis, the company can create new markets, potentate customer fidelity and increase the effectiveness of the company.
- Decrease costs and time:

Implementing Business Intelligence within a company reduces the high cost of human labor when it seeks to obtain answers without an appropriate tool, minimizes the costs of errors and miss-communications. More accurate data means better decisions and better decisions always save company money. Reduce operating costs because standardized and trustworthy data can limit redundant processes and departments, it's easy to see how long it takes departments to get into loops , find and collect data that should not take so long, waste time that could have been spent doing something far more productive, the monitoring capability of Business Intelligence frameworks allows process administrators to identify problems and new market opportunities. All this, combined with reports and analysis, can be a powerful tool for companies to avoid unnecessary costs and focus on important areas. Taking into consideration the human errors when manipulating data with the implementation of BI, make it possible to initiate a process starting with better data and ending up with better decisions

- Decrease risks: Analytic a fundamental process when it is about taking risks because at the end it is possible to effectuate some prognostic's that with real and trusty data is possible to prevent from dangerous investments and to increment efforts in some of the others products that are part of the company project portfolio.
- Increase Data Governance: There must be a unique governance group that should provide some data quality process where it could be possible to provide the different departments of data that will be unique and approved, in order to do all the sub sequential activities.

If the objective of an organization is to change the culture in order to transform it in a company data- driven then Business intelligence is a strategic framework in order to support and help to the decisions of the business. Is consider one of the main methods that allows managers an employees to transform un-structured data on information and facts useful r to achieve company's vision Is a method that obviously, with other agility methods will allow to forecast and to aligned the projects and process with the strategy and goals. As porter said one of the main basics in order to set an strategy is to identify how to set a competitive advantage, and taking into account the innovation curves and the constant mobility to offer better services and products to the final

customer the incumbents in the markets have to set a lot of efforts in order to search and acquires competences and capabilities, that's automatically will convert a company in a flexible one, is important that all the research should be aligned with the company strategy.

Bi implementation in a company provides tools to improve productivity and reduces the time spent on manual tasks, or traditional analysis that doesn't allow the user to go further ; it also allows managers to analyze each process separately, by showing them the conditions of it at any time as well as helping with accounting . Moreover, the organization acquires the capability of reacting faster to change when there are necessary changes to produce on the processes.

The BI implementation in a company provides tools to improve productivity and reduce the time spent on manual activities, or a traditional analysis that does not allow the user to go further; it also allows managers to analyze individually the processes and also then to get a general conclusion taking into account the relationship between processes , showing them the conditions at any time and helping with accounting. Furthermore, the organization acquires the ability to react when a critical situation is presented and then it is necessary to create a change within the company.

Allowing the automation of processes by involving the combination of human activities and information technology applications. Before, the golden rule was: "First organize, then computerize" [32].

3.3.1 Reporting

What is considered as one of the Best practices according also with the ITIL is to gather the data about the facts that already happened inside a company for example in some retail companies is relevant for the different departments to identify some keys of successfully like how many products of a certain brand were sell, to know if a certain promotional campaign was successful or not. To understand what was the customer driver in order to take a given purchasing decision. In order to allow the human resources to have access to this data is important to give as a support some documents that will give visibility of this type of information. Reporting is a fundamental part, to move forward the business intelligence framework just starting from the data governance, to have access to this data in order to give some quick wins for the next steps,Often implementation of the reporting involves extract, transform, and load (ETL) argument treated during the subsection (3.3.2) depending on the system that is gathering the data for the company. *DWH*. The reports can be presented as dashboards as can be seen on the figure:

CHAPTER 3. THEORETICAL FRAMEWORK



FIGURE 3.8. Abax Bi-Metrix, Example of report. Figure reproduced from [1]

Usually, these reports are also made according with the needs of the given departments and the are established when the requirements are specified with a given format, are a standard base under which everyone creates the work logic and use different other tools to allow this statics reports to give some analysis and to speak more. The reports are based on some metrics and parameters and are delivered also following a certain business logic. The previous definition is how reporting is considered traditionally within the business but when considering BI reporting also includes, moreover, than just presenting data also information this, in a deductible way means that, there is an analysis process included. So the Final user will be in the capacity of identify the facts and act upon them.

3.3.1.1 Storytelling

It is important to clarify that the purpose of the graphs and visualizations is to tell the strongest story, the purpose of creating a visual is to convey a message clear, so it is really important to study who to transmit this information. It is also a matter of agility, it is how to get the story across the viewer; for this aim there are many ways, charts and graphics . the importance is to select the right one in order to evidence the right information. Each of them are made in order to achieve a particular task.



FIGURE 3.9. Choosing a good visualization. Figure reproduced from [2]

Is really impressive how the work of Nathan Yau made viral as how he say's

I can tell you that about 40 percent of people age 25 to 34 are working on an average day at three in the afternoon. I can tell you similar numbers for housework, leisure, travel, and other things. It,Äôs an overview. What I really want to see is closer to the individual and a more granular sense of how each person contributes to the patterns. I want to see how a person,Äôs entire day plays out. (As someone who works from home, I,Äôm always interested in what,Äôs on the other side.) [24].

What he did was to look into micro data where was evidenced what American people used to do in a normal day in order to construct the next storytelling



FIGURE 3.10. Example of StoryTelling. Figure reproduced from [24]

3.3.2 ETL and ELT

ETL AND **ELT** are two different approaches by which is established the procedure of how the information from different sources is gather inside a DWH within a company. what really change is not just the order of the operation but first each of the letters stands different procedures that are applied to the data:

- Extraction: When data inside an unstructured pool is migrated to a temporary area of stage.
- Transformation: Process in which the data is transformed and structured in order to enter with the format required inside the DWH system.
- Loading: When the data from the repository is moved to the DWH to effectuate the analysis.

As was described before is not just a matter of sorting the process, when applying *ETL* approach the final user has to wait til the data has completed the whole process it also includes the fact of dealing with the capacity machine, this traditional approach introduce the fact of analyzing which type of raw data should be converted and analyzed before the loading, but what happen if the company after a certain period of time realizes that should make some analysis about information that first wasn't considered as critical. when applying ETL approach is not possible to recover this information.

Because of that, currently, there are so many studies in the market in order to allow a new process in which is possible to gather the raw data without doing the transformation first,

In this case is the *ELT* As the name describes the process of loading data is made before the transformation process. Both approaches must be analyzed taking into account the company vision, strategy, the experience and costs.

3.3.3 BI Components: OLAP,

OLAP stands for On-Line Analytical Processing, it performs a multidimensional analysis providing the capability of managing the data to do different calculations or to be displayed in different ways, OLAP enables end-users to perform ad hoc analysis of data, being a powerful system of data discovery ,including capabilities for limitless report viewing, complex analytically calculations, and forecast about possible scenarios. OLAP considers each single attribute as a dimension different from data relational; OLAP in order to do all the operations and consolidate the data should make some processes as:

- 1. Consolidation: Including the aggregation of data according to customers request.
- 2. Drill-Down:It is a technique allowing to navigate through the information.
- 3. Slicing and dicing: To manipulate the data based on the attributes set up.

3.3.4 BI Components: Data Discovery

Data Discovery is intended for the process than involves the continual research of answers about a certain amount of data, it is really important because it works to find anomalies, criticities, predicting the possible outputs. Is a tool that allows the users to predict and even with the decisions to take in order to increase the profits and change the course of the whole enterprise. The process of discovering not expected connections or hidden information, is related to more disciplines, because with the research and continue advances it is possible to make it in an easy way, eliminating the tedious and manual work ; statistics (use of quantified models in order to describe certain behavior), artificial intelligence (what is possible to reply from humans rational behaviors to a machine) and machine learning (The machine through data can do some predictions), allows to get some insights from price to promotions or even customer behaviors and to avoid possible risks.

Chapter 4

Finiper Group

During this chapter, it is described all the information related to the company Finiper Group where this dissertation is being carried out. It is a recognized company in Italy and for this reason, it is worth highlighting the way the company has develop his processes. In first instance, it is presented an overview of the company during the section (4.1), followed by the merchandise structure, Description of the IT department (4.3), then a brief description of its products section (4.2) to finish then with an specification of the thesis processes that are going to be updated during the development of this dissertation.

In the current competitive markets of the retail domain where there is a constant change on the customer expectations and the services offered are growing exponentially all come along with a fast technological development, the firms are in constant search of opportunities to reduce costs and to increase profits, seeking a revenue that exceeds considerably the costs of production, with the objective to convert the strategy more profitable.

Iper Montebello Corporation operates around Italy in about 7 different regions. It has approximately 8.000 employees that work in all type of areas, such as design, research and development, product development, manufacturing and sales. It is a company that is characterized for being highly client oriented, always looking for better solutions in order to improve customers experience on each of the 26 stores located around Italy. Iper's Company vision is to be both, quality and accessibility, turning all the stores into the ideal places in order to do shopping. Helping people in order to live better .Iper as part of the finiper group has a business logic that operates in collaboration with different business areas.

The thesis work is executed in Iper Montebello a company of the Finiper s.p.a which is ruled by the disciplinary mechanisms imposed by the control agents reflected on the (4.1)business unit, which is located at Milano in Lombardy, Italy. The current study is done in the Information systems department, therefore the company description is now focused on the characteristics and data of Digital innovation team inside the department.

4.1 Organizational Description

Along this section, are illustrated some characteristics of Finiper as a company, this is something to focus on, given that in the past few years it has been growing strongly and now it has become a recognized company in the retail-trade domain . First it is presented a brief description about the company (4.1.1), then the company's mission (4.1.2) and vision (4.1.3). The information exposed along this section has been taken from [15].



FIGURE 4.1. Company control agents. Figure reproduced from [15]

4.1.1 About the company

Iper Montebello is a company with more than 40 years on the market and strives in order to

empower customers decisions related to the daily consumption, the philosophy of the company is centralized in giving customers the possibility of having a instantaneously meal, allowing the client to have knowledge about the process and final product. *Iper Montebello* is a company founded under the principles of a *GDO* company (Grande Distribuzione Organizata), and includes a modern system to sell throughout a supermarket network, in this particular case the aim is to represent the development of the traditional stores.

Iper Montebello operates under the Finiper Group in a large- scale retail trade sector that will increasingly require high-tech and innovative IT solutions to increase the efficiency of the operations and in their own (back-office). Finiper Group is divided into three large areas: hypermarkets with the Iper, La Grande i, UNES supermarkets and real estate businesses.

- **Hypermarkets:** The hypermarkets of Iper, La Grande i are present in 7 Italian regions: 27 points of sale (4.2), distinguished by the high quality of fresh products, the wide range of food and not-food, in addition to the extreme care of the sales areas. The wide range of regional specialties, branded products and the assortment of branded products are essential traits.
- **Supermarkets:** UNES supermarkets are present in Lombardy, Piedmont, Emilia, Liguria and are subdivided in the following categories:
 - 1. **U1** As you want me: these are traditional supermarkets with a strong relationship with the customer, favored by the rooting on the territory and by the size, which never exceed 1,600 square meters.

4.1. ORGANIZATIONAL DESCRIPTION



FIGURE 4.2. Store Map and Markets, November 2017. Figure reproduced from [15]

- 2. **U2** Supermarket: with their commercial formula they combine the quality of Unes supermarkets with a stable and advantageous price line.
- **Real state activity:** The real estate business of the Finiper Group consists in the management of the commercial galleries distributed in northern Italy and along the Adriatic backbone. To support the services of the Finiper Group, there are also strategic activities such as the management of logistics platforms for the supply of raw materials.

The origins of the company take place in the 70's, when the entrepreneur Marco Brunelli had a vision about the importance that the hypermarkets would take in the Italian industry, decided to build a whole new business unit . After some time, the business model turns into a successful activity represented by its incomes. The products were recognized by the consumers because of it's quality and the demand was so high that it was not possible to meet it through just some stores, for these reasons in the following years, the company *"Iper la grande i"* inaugurated more than 10 stores, and the business increased the level of production and workers.

Also in the same decade, Finiper was founded, which will become the Group's holding company. "The visionary ability of this entrepreneur feeds the virtuous circle that leads from innovation to growth, to new research and experimentation."[15]

Twenty years later, after the foundation of Iper Montebello, in 2004, the metropolitan hypermarket will be conceived: Iper Portello, in the center of Milan, a brand-new store which will contribute to the redevelopment of the district, obtaining particular recognition from the Building Commission of the Municipality of Milan for the urban and architectural quality. The growth continues and in 2016 the 27th hypermarket Iper, opened in Arese, was inaugurated, a unique excellence in terms of quantity and quality of services offered in the Milanese belt.

4.1.2 Mission

It is important to have into account that Iper Montebello is a customer centered company who is maintaining his tactical position and improving in different areas in order to survive in the business world, the problem, comes later when the company starts to manage a huge amount of data and starts to run behind targets without establishing a clear common horizon. So the components of the organization begin to spin under the uncertainty, Making inevitable the fact of creating individual parameters and guidelines affecting the methodology, the communication between different departments and a business analysis approach. Considering this situation,

"The more data-rich your business becomes, the more important it is to ask the right questions at the beginning of the analytic process. That's because the very scale of the data makes it easy to lose your way or become trapped in endless rounds of analysis." [22].



FIGURE 4.3. Holding Organization Chart . Figure reproduced from [15]

The goal, of the organizations is to find a good balance between demand and supply, summarized in: to optimize the efficiency of the company and to increase its attractiveness. Some possible consequences that may occur if this balance between demand and supply is not reached are: loss of sales, decreased delivery capacity, increased stocks and higher unit production costs among others.

4.1.3 Market focus

In the last years with the e-commerce explosion, the clients have been habituated to change stores during the purchasing process always of course, waiting for a continuing and pleasant experience, this behavior involved an omnichannel strategy from different manufacturers and retailers form the consumption sector, in particular for the logistics. that's why there are five objectives which are the drivers of the institutional strategy.

- 1. Always to preserve the interest of the consumer in all the phases involving the selection and production processes.
- 2. Nutrition culture diffusion and conscious consumption.
- 3. Selection and continuous research for a eco-compatible packaging.
- Reduction to the minimum the environmental impact of the activities developed during all the core business processes.
- 5. Increase the amount of solidarity initiatives including a local and global perimeter.

From the first day, Iper has a single objective: to select, where possible, 100 Percent Italian products, with scrupulous attention to fresh, small local productions and traditions at risk of extinction. The producers, farmers and breeders with whom the company has established a direct collaboration, and who share the same values to guarantee quality, traceability and sustainability.

This means: Always prefer Italian products wherever possible, when selecting suppliers and in the design of the private brand products. Being able to innovate by responding to the needs of a changing country. Integrate the national assortments with local products, coming from territories suppliers where our stores are located, to the advantage of local economies, of freshness of products and environmental protection. Take care of every detail in the design of points of sale and in the display of products, creating a relaxed and pleasant atmosphere that refers to the Italian squares and markets.

4.2 Products

Iper Montebello has a set of products offered to the customers, inside the large-scale retail domain. These are all contained into **Food and No-Food** and here they are grouped into a determined merchandise structure depending on their orientation and purpose. The company has develop a wide range of food products and other different type through a diversification on own brands or exclusive ones, is one of the answers to the requirement of offering safe products, with high quality standards and with competitive prices, the offer variety is aligned with the market changes and the new customer needs.



FIGURE 4.4. Iper la grande, different brands. Figure reproduced from [15]

The decades of experience in the retail industry and the indisputable quality of its products have made *"Iper la grande i"* a model of enterprise to follow, capable of giving value to its brands and allowing it to grow in the new markets. The company has around 900.000 products to offer in the stores located all around Italy, those Products are classified under a given codify systems as *codici interni*,due to it's complexity, Iper has established a certain logic in order to control and develop different analysis.



FIGURE 4.5. Iper la grande, Merchandise structure. Figure reproduced from [15]

All the information presented during this section has been reproduced from Iper Montebello internal documentation and the Iper web page. [15]

4.3 Information Technology Department

Along the different departments dedicated to the management of business processes, the Iper IT department, denominated D.S.I (Digital Sistems Informations) localized in Milan - Italy, has the task of provides opportune and effective input information for the other functions and areas of the company (Marketing, Sales, Production, Logistic, Planning, Distribution, Financials...) for all markets in the different localization's among the country.



FIGURE 4.6. Location Iper. Figure reproduced from [15]

During the last years, the principal challenge of the different departments inside of Iper la grande has been the lack of data governance, definition of standard processes and tools with the objective to guarantee the faster communication, preserving the information between the numerous realities or areas of the national network.

The aim of Iper's executive departments is therefore to create common and standardized procedures that will improve synergies between the components of the different functional areas, with the finality to increase the overall efficiency of the entire organization. Over the years Iper has identified the necessity of standardize and to centralize the processes, data and facts, even to achieve economies of scale. For these reasons, the DSI department dedicated to the implementation and maintenance of the IT platforms, that support the processes of all Iper group, from the creation of mail accounts, token of accesses, connectivity, hardware, functionality, infrastructure and others and which allow them to have an information structure that is up to the present industrial organization. In summary, the IT department has different functions: Initially, it guarantees seamlessly the flow of information that characterizes a complex and articulated scenario such the Iper Montebello company, with particular attention to the issue of segregation of sensitive content; On the other hand, it offers simple and intuitive IT solutions for those who are working with large amounts of data daily, or for who has some difficulties with the platforms, implementing and customizing the capabilities provided by the information system **Microstrategy** and others systems. The local IT has the direct contact with suppliers of different systems and guaranty the correct assistance to the Iper users of all the markets.



FIGURE 4.7. E-commerce Architecture . Figure reproduced from [15]

Chapter 5

As-Is Architecture

This chapter is intended to illustrate the tools of analysis used by the company for many years, this is to open up the discussion of the **AS-IS situation** of Iper Montebello before the implementation of a Business Intelligence Framework plan execution (situation later analyzed inside the Implementation dedicated chapter (7)). First of all, it is presented the concept of Microstrategy, subsection (5.1.4), as the standard tool within the company to obtain the data known inside the DWH for the other departments; The concepts of CercaCodice, Webgate and Twittiper on the subsections (5.1.1, 5.1.3, 5.1.2) as the business process management tools for analyzing data, from which all the main decisions are based on. Moreover, it is also presented the role of the different users for the design and definition of the processes of the company in specific the role of the DSI department and the process about the requirements.

5.1 Architecture As-Is

In this section, it is described the current situation of the Architecture in terms of source of information. The Analysis of the AS-IS situation is necessary because is the starting point to establish an adequate, coherent system of reporting and analysis of the data inside Iper Montebello Company. For instance, the metrics that are delivered by the DSI department to the rest of the company using data from the DataWarehouse must correspond to the Logic's and expected data that are needed in order to take crucial decisions within the company, To illustrated some of the arguments to take into account are: quantity to buy from the providers, alliances to make with other companies, benefits of promotional campaigns and so on. Once the AS-IS situation is modeled, the criticalities found on the current state of the processes and governance of the data are exposed. The criticalities are mentioned based on the flaws on process mapping, metrics created, information system structure and data availability to the other areas of Iper. Subsequently, the establishment of the AS-IS situation a new proposal is presented in order to improve the processes within Iper Montebello company. And it is established following the theoretical approaches developed in the chapter (6).

First of all, it is important to specified the type of architecture that the company is following in order to obtain the data, to structure it and finally to deliver it to the end- user, which are the ones in charge to make assumptions, predictions and moderate the data. In the figure (5.1) it is possible to evidenced the As-Is situation of the company Iper Montebello System. The DataWarehouse works with an Oracle DataBase structure SQL developer but all the information that is collected in it uses a logic of migration from other sources based on ETL as has been described for the purposes of this dissertation during the chapter (3), that means that the data after passing trough and ODBC System provided by Oracle is already selected and is not raw data anymore, is possible to interrogate it and structure it so the Iper DWH can read it,



FIGURE 5.1. Architecture DSI. Figure reproduced from [15]

On the figure 5.1 is possible to evidenced all the systems that are daily feeding the DataWarehouse; each of them or almost all provide information about the profits, the logistics, the purchasing and sales that where done with a certain historicity:

- NSDR is a system that provides information to the DataWareHouse about the sales that where done in a point of sale with a detailed description about the three of the main dimensions established within the company that are:
 - Product
 - Ente (Place, Point of sale)
 - Period (YYYYMMDD in which the transaction was made).

This system also provides information about the assortment, the providers of the products, the inventory and the transactions that are made inside each sale point or between the sales point, the product prices, promotions and so on. all fundamental data that is involved by means of a supermarket, and in order to take different decisions is used by the purchasing department, sales , management control and so on.

- Open Retail: Is the system in which are codified all the products and are integrated then in to the registry, with their respective attributes as Codes, data and given Merchandise structure, that ifor the specific cases of Iper Montebello are made under the following order:
 - 1. Department
 - 2. Sector
 - 3. Group
 - 4. Family
 - 5. SubFamily

In specific this could be appreciate on the figure 4.5

- IperDrive: Is all the information related to the sell and transactions made for the IperDrive service within Montebello Company, IperDrive is the service in which is possible to order online and the user can go to pick it up already done directly in a sales point selected.
- Asar : Information related to the cashier system.
- SIGO : All the information regarding to the Ortofin structure
- GOLD : Back office of the logistic platform.
- CLL :Loyalty system.

As was defined on the figure (5.1) there are different applications that are in charge of interrogating the different systems, including interrogations not standardized and neither established under a given and unique process, for the reasoning of this chapter each of the system that are used by the different areas of the organization will be enlisted as follows:

- DSI . Direzione Sistemi Informativi
- Acquisti: purchasing.
- Marketing
- Controllo di gestione. Management Control.
- Human Resources.
- Personnel Administration.
- Vendite. Sales

The dissertation will only focus it's efforts on the purchasing, sales, management control and marketing requirements.

Before of defining the applications used by the different users it is important to understand the logic by which the *DWH* is based: The Iper DWH is structured by the logic of relational Tables as was already explained during the chapter (2) in the subsection (3.2.1.1) and is constructed taking in to account a back up system in case of a necessity. The main tables are classified by:

• Registry Tables: in the registry tables are registered the dimensions and it's given attributes as:

 $(5.1) Dim_P rod$

In the (5.1) it is possible to identify different fields in which are some attributes that where already selected transformed and loaded in order to be displayed to the end user as an example of the relation tables there is the DimForn displayed in the fig

$$(5.2) Dim_Ente$$

5.2 Is the table of the registry of all the point of sales of Iper Montebello among the italian country.

5.3 Is the table in which is specified the measurement units of the calendar with the days and the format.

(5.4)

Dim_Forn

COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
1 FORN_ID	NUMBER	No	(null)	1	(null)
2 FORN_CD	VARCHAR2 (12 BYTE)	No	(null)	2	(null)
3 FORN_DESC	VARCHAR2(100 B	No	(null)	3	(null)
4 FORN_PIVA_CD	VARCHAR2 (50 BYTE)	No	(null)	4	(null)
5 FORN_DATA_INI_ID	NUMBER	No	(null)	5	(null)
6 CARREFOUR_CD	VARCHAR2 (13 BYTE)	No	(null)	6	(null)
7 DATA_REPLICA_ID	NUMBER(13,0)	Yes	(null)	7	(null)
8 INSERIM_DATA_ID	NUMBER	Yes	(null)	8	(null)
9 CONTR_ID	NUMBER	No	(null)	9	(null)
10 AGGIORN_DATA_ID	NUMBER	Yes	(null)	10	(null)
11 FOAM_CD	VARCHAR2 (8 BYTE)	Yes	(null)	11	(null)
12 FOAM_DESC	VARCHAR2 (60 BYTE)	Yes	(null)	12	(null)
13 COND_PAG_CD	VARCHAR2 (4 BYTE)	Yes	(null)	13	(null)
14 PRES_LISTINO_FLG	CHAR (1 BYTE)	Yes	(null)	14	(null)
15 BLOCCO_CD	VARCHAR2 (1 BYTE)	Yes	(null)	15	(null)
16 BLOCCO_DESC	VARCHAR2 (50 BYTE)	Yes	(null)	16	(null)
17 CONTR_CD	VARCHAR2 (12 BYTE)	Yes	(null)	17	(null)

FIGURE 5.2. Structure Table DimForn. Figure reproduced from [15]

• Facts Tables: in the facts tables are registered and historicity all the events that has occurred as sales, transactions on the assortment in specific the following enlisted:

$$(5.5) $DimPrezziPdv$$$

5.5 Table in which is gather all the data related to the prices of the products according with the point of sale.

(5.6) FactStockGiorno

5.6 in this table is located the information of the sotck in each Point of sale.

5.7 Information related to the purchasing transactions.

(5.8) Fact Ven Tot Giorno and Fact Ven Promo Giorno

5.8 in this table is possible to find the data about the sale of every date by point of sale and by product.

(5.9) Dim Tst Promo and Dim A sj Promo zione

5.9 in this table all the promotional information by product and point of sale taking into account the date is presented.

Furthermore, it will be a deeply understanding of these tables and how they are fundamental in order to understand final user requirements, the data that is really available within this machine of 10 terabytes and the occupied space by each of them.

5.1.1 CercaCodice

This is a tool developed by one of the employees is not standard and should be evidenced that there is not a governance of the data, because the application was developed by the use of macro VBA in excel that interrogates directly other systems without passing first through the Data warehouse. In this way the process of data quality is affected and not guaranteed. Being an excel instrument, assumes also the use of VlookUps and the increase of the data needed for the analysis every day is evidenced by all the problematic using this system. The use of this system requires to many people managing the data inside it and the size of the information starts to be a problem, The detailed information in an Excel spreadsheet typically does not give a quick update. Excel is really limited detailed and analyzing detailed, key steps in the process can be missed and delays in one step can push out the completion of the project. The application in excel calls the information directly from the sources throughout a query system ODBC as specified on the fig 5.3:

	Source of information	Data Extracted	Used By:	Number of people stimated
C e	DWH	Anagrafica Referenza	Buyer	
r c a C o d i c	SDC (Open Retail)	Vendite/Acquisti	Pricer	50
e	GOLDEN	Fornitori	Codification Office	

FIGURE 5.3. Cercacodice Main Characteristics . Figure reproduced from [15]

After doing some analysis inside the company in order to determine if the tool is fundamental to make given analysis and to develop daily operations. it was determined the percentage of the company that utilize the tool from a sample of 39 people: and according with the scale moreover the 50 percent considers this tool as indispensable:



FIGURE 5.4. Scale of the survey applied. Figure reproduced from [15]



FIGURE 5.5. Level of utilization CercaCodice. Figure reproduced from [15]

The objectives were explicitly defined by the company in terms of perimeter when the application was first created but then with the increase of products and point of sales started to be considered a bit hard to manage . However, these objectives were just referred to the departments involved but not taking into account a governance process of data. In specific this application allows the user to have access to different information as can be seen on the figure 5.6 .There are 3 possible option of research: Estrattore, Cifra , Lista, Codifica, the first one is the base one and allows to extract massive data information by interrogating the systems according to the references of products;cifra, allows to compare based on the references selected the sales

amount of the current year with the previous one; Listato, allows to identified all the attributes of the products selected, and last but not least Codifica, interrogates the registry. listed bellow are the most common activities to be developed according with the buyers and the prizers:

- It is a really useful system when fast research about references in meetings with providers.
- Information about reference purchasing price, sales and logistic about the information that is acquired by the platform, discounts and so on.
- Report about the daily sales.
- When monitoring processes are included this application is really use full specially referred to sale and stock to identify some alerts in the process.
- Daily researches about using the screen of extraction to verify the new charges.
- Analysis of the sales according to the product structure.

E Se de s			WR-CECA	CODICE_IB2017.eleb (Sola lemura) - 1	iacel	8	Massianiano 80 - 0 X
Norma Tradict Laplors page Image: Straps Segne UI -8 Image: Straps 6 C 2 10 Image: Straps 6 C 2 10 10 Image: Straps 6 C 2 10 10 10 Image: Straps 1 2 4 10 1		Image: Second	0 V Control Victor Vic	Tholo 42 Valore valido 2	Tetal Tetalo 2 Warning Text Normale	Natoria non v	S terrore substrations - Area -
	C D RESET REPORT	E F G F	I J K L M	N O P Q	RST	U V W X Y Z	AA AB AC AD /-
	5		n Pecier Texture Constraints				
2 Color Tapler field feather 5 Color Tapler feather 5 Color Tapler 5 Color Tapler		8,000 0,88 8,00% 9,000 0,88 8,00%	Codes Alteren 6 2017	500 538 500 539	0 0		

FIGURE 5.6. Cercacodice Application. Figure reproduced from [15]

5.1.2 WebGate

Webgate400 is an application provided by IBM installed on AS/400 and promises to improve and speed up the daily work of the users and to optimize business operations. But the management costs are increasing, at this moment are more than 100 thousand per year, The system isn't in a continue improvement because of the investments that this operations require. There are also some bugs that were already identified by the users and made this tool useful just for some areas

within the company with some analysis in specific as can be appreciated on the fig (5.7), the research can be done just by selecting the provider or the reference. The data flow are send to the machine of the *WebGate*.

	Source of information	Data Extracted	Used By:	Number of people stimated
	DWH	Anagrafica Referenza	Buyer	
W e G a	SDC (Open Retail)	Vendite/Acquisti		30
t e	NSDR	Stock Warehouse/Platform		
	Gold	Competitors		
		Fornitori		

FIGURE 5.7. Webgate Main Characteristics. Figure reproduced from [15]

The problem of research perimeter limited is evidenced on the number of users and areas within the company that considers this application as key to develop the daily tasks. Following the scale represented on the figure (5.4) this information is evidenced on the figure bellow (5.8 where from a sample of 39 users just the 23 percent consider it essential but more than the half of the sample considers it useless.



FIGURE 5.8. Webgate Level of utilization. Figure reproduced from [15]

Although the quantity of users is not really high, the few people that uses it describes the tool as

the air, essential to subsist; because this is a tool that allows the user to find answers in a short amount of time while some confronts with the providers are done. The users highlight that is always about to minimize the time of research information like:

- Information based on references or providers among the current year, and its comparison with the previous year.
- Current Promotional Campaigns and a historicity not longer than two years.
- Information from the encoding.
- Price of Sale (no historicity).
- Cost of purchasing from the last entered.
- Stock situation.
- Delivered on the current year made by the providers.
- Competitors.

r Usta Retet Anggalica Precedente Jogo (FS) (FS) (FS) (FS) (FS) (FS) (FS) (FS)		
WEBGATE400 Session		
RICERCA Articolo Descr. cliente: Barcode Fornitore Outi Anticolo Stato Iclasse assort. : Tipo assort. : Tipo prezo : Variato Variato Harchio :	FUNZIONI D. STUME.referenza R. POV (F14) COUND (F16) COUND (F16) SETTIMANA (F17) SETTIMANA (F17) Consolidato Consolidato Consolidato Consolidato Consolidato Consolidato	Listini acquisto (roz) (roz) prodinari (ris) report Acquisti for contracti Contracti Contracti (user field) (user field) (roz) (user field) (roz) (roz) (roz)
-RAGUIDENENTO MERCEOLOGICO Reperto : Settore : Gruppo : Faniglia : Sottofamiglia :	piattaforma (96) On order Ricevimento merce (F11) /assegnaz. (F13)	Crazi a carre Crazi Articoli del Articoli del Menter (roz) Menter (roz)

FIGURE 5.9. Webgate Application. Figure reproduced from [15]

5.1.3 Twittiper

The system is made up of excel sheets with a power pivot plug-in for the easy consultation of data coming from the data warehouse system, the sheets are consulted via rdp or vmware

horizon access from different devices, thin clients, notebooks, tablets. For *vdi* is intended virtual desktop infrastructure.

Currently there are 89 vdi-marketing in production, which are headed by a vdi called vdmarketing-00.

Vdmarketing-00 exposes a share through which the other vdis access the excel sheets, sharing is write protect there is no possibility to make changes by users Vdmarketing-00 exposes an ftp folder, this folder is populated with exports sent by dwh with weekly and monthly data, the monthly mailings take place on the 1st of the month and contain the data of the previous month, the weekly mailings take place on the night between Sunday and on Monday and contain data from the previous week.

The data is refreshed in excel sheets by the scheduled execution of two scripts from vdi vdmarketing-00 and stored in a given path Scripts, if necessary there is the possibility to execute them manually.

When accessing the users in the respective vdi a script is performed through group policy that takes care of deleting all the files and folders in its desktop and copying two folders with links to excel files from vdmarketing-00.

There are different issues that have been identified:

- 1. There is the possibility of having false positives and misinterpretation of some fields.
- 2. Sometimes the vdi are locked and are not accessible, they need a manual reset.
- 3. Limit of 20 sessions with simultaneous access to the share exposed by vdmarketing-xx.
- 4. Sometimes the data send stream in FTP has crashed.
- 5. When a rdp / horizon session crashes from a vdi the temporary files generated by excel are not automatically deleted, and in case of repeated drops of the session the vdi becomes unusable due to the saturation of the disk space.
- 6. The script that automatically deletes the folders on the vdi desktop does not work if there is something in the folders that created the user and does not come from the automatic copy.

	Source of information	Data Extracted	Used By:	Number of people stimated
T W i t	DWH	Trends comparing with the previous year	Manager of the departments on each point of sales	98
i p e r		Sales montly and weekly	Marketing	

FIGURE 5.10. Twittiper Main Characteristics. Figure reproduced from [15]

Twittiper is not an enterprise tool but although it has some benefits. the difficulty arrives when the user needs to make data discovery and the fact that the maintenance moreover than expensive is really difficult to develop and the utilization cost are between 60 thousand just for the 98 virtual desktop infrastructures.



FIGURE 5.11. TwittiperLevel of utilization. Figure reproduced from [15]

The Application was made by one of the employees and is possible to effectuate the access through a virtual machine. This application compared with the others has the advantage of taking the information directly from Iper DWH as is represented on the figure (5.1)this process makes possible to increment the data quality process.

HOME INSERISCI DAYOUT DIS	AGINGA I	ORMULE	DIAL	11 1425	13JOINE	VISU	AL LEZA	POWE	Phot A	MALIZZA	PROGETTAZIO	3745			
▼ 1 × √ fr 455	,7														
8	D	E		F	н	1	1		J	ι	м	N	P	Q	R
	*	Scegli u	ina o pi	iù settim	ane			7	Reparts	Ð			×	_	
1 Area 2 Area 3 Area		01	02	03	04	05	06	07	4					Reset	t
	_	08	09	10	11	12	13	14					_	_	
	<u>7</u>	15	16	17	18	19	20	21	Settore		_		<u>×</u>		
intebello 02 Cremona	^	22	23	24	25	26	27	28	30	31	32	33			
embate 04 Magenta		29	30	31	32	33	34	35	34	35	36	37			
lbiate 06 ∨arese		36	37	38	29	40	41	42	39	45	83				
riate 08 Rozzano		43	44	45	45	47	48	49				_			
bicone 10 Busnago		*0		10	40		40	47	'						
stelfranco 14 Udine	×	50	54	52	33										
	Tren	d % Fa	ttura	ito		Tre	nd Fa	tturat		Tren	ıd Pezzi/	Kg	Evoluz	ione Pres	sione
												-		Promo	
io/Famiglia 💌	015vs2014	2016vs20	15 201	7	2015v+2	014 2	016vs20	15 20	7	2015vs2014_2	016vs2015 2	017	2015vi2014	2016vs2015 2	017
ICOLI CUCINA/SACCH.ALIMENTI	60.42%	10,64	175 196	-10,98%		5.120		5.357 8.783	-€ 8.616	49.45%	91,73%	-14,89%	15,7%	-6,5%	-1,3%
ICOLI USA E GETTA/MELAMINA	-11,35%	-2,10	1%	-14,23%	-6	13,407	-6	2.199	-€ 14.596	-4,70%	-2,76%	-3,07%	-9,1%	8,6%	47,8%
TER.PENTOLE/CAFFETT/THERMOS	-15,11%	-8,64	1%	-6,57%	-63	10.939	-6	5.310	-€ 3.686	-10,53%	0,47%	-10,54%	-6,1%	6,0%	3,5%
CHIERI/BOTTIGLIE/BROCCHE	-14,72%	-17,14	6%	-9,87%	-4	4.194	-€	4.165	-€ 1.988	2,17%	-10,98%	-19,07%	10,1%	-6,1%	1,5%
IDEL/INCENS/ESSENZ/POT-POURE	141,61%	-11,31	1%	-14,57%	(4.556		€ 879	-€ 1.004	179,78%	-23,58%	-15,40%	32,4%	-15,5%	9,0%
T FORMO (FTAND (TEC) (BOT AUD	7.0656	19,3	379 296	-29,83%		1 916		1.150	-€ 4.819	20,30%	8 13%	-22,62%	-5,3%	-12,3%	7.4%
IT.FORNO/STAMP/TEGL/ROT.ALIP		-16.1	1%	1.31%		-€ 876		€ 651	6.44	-11.61%	-24,50%	1.07%	-1.9%	1.4%	-7.4%
IT.FORNO/STAMP/TEGL/ROT.ALI# TI/INSAL/TAZZE/TEIERE/ECC. GIONALI/TECN.GEST/OBSOL	-17,81%		104	-26,76%		1.775	•	5.542	-€ 12.203	5,56%	17,91%	-24,87%	16,9%	-4,3%	-6,0%
HT.FORNO/STAMP/TEGL/ROT.ALIP TTI/INSAL/TAZZE/TEIERE/ECC. GIONALI/TECN.GEST/OBSOL. NSILI CUCINA	-17,81% 4,63%	13,83	3/0				6	5.943	-€ 54.536	2,04%	3,12%	-12,45%	-0,1%	1,2%	25,2%
IT.FORNO/STAMP/TEGL/ROT.ALI/ TTI/INSAL/TAZZE/TEIERE/ECC. GIONALI/TECN.GEST/OBSOL NSILI CUCINA complessivo	-17,81% 4,63% -4,42%	13,8	2%	-16,36%	-6	15.154									
IT.FORNO/STAMP/TEGL/ROT.ALIP TTI/INSAL/TAZZE/TDERE/ECC. GIONALI/TECN.GEST/OBSOL. NSIELCUCINA complessivo	-17,81% 4,63% -4,42%	13,8 1,8	2%	-16,36%	-6.	15.154									
AT.FORNO/STAMP/TEGL/ROT.ALIP FTI/INSAL/TAZZE/TEERE/ECC. GIONALI/TECN.GEST/OBSOL NISELCUCINA romplessivo	-17,81% 4,63% -4,42%	13,8 1,87	2%	-16,35%	-E	15.154									

FIGURE 5.12. Twittiper Application. Figure reproduced from [15]

For the previous systems there is not documentation available where can be identified the source of the information and the processes applied to it.

5.1.4 Microstrategy

Following there are listed some of the funcionalities that are allowed for the users when using Microstrategy and is completed with the figure (5.13):

- 1. Data source: Company Data Warehouse (DWH), This is the most important characteristic where the data is loaded directly from the dwh tables.
- 2. Extrapolation of "certified" data from the company dwh in the form of reports (and dashboards).
- 3. Execution of pre-set reports ad-hoc for users and they have the possibility to set up reports according to the level of detail desired.
- 4. The user can access pre-set reports on MicroStrategy by navigating the data through OLAP operations (drill-down, roll-up etc ...).
- 5. Used for reporting, One of the main problems from this solution is that the on-fly search for a data is not immediate.
- 6. The user can receive pre-set reports in Excel format via email.

М	Source of information	Data Extracted	Used By:	Number of people stimated
i				1 () () () () () () () () () (
с				
r				
0				
S			All the departments	
t	514/11	All the information		A.H
r	DWH	available on the DWH		All the organization
а				
t				
е				
g				
У				

FIGURE 5.13. Microstrategy Main Characteristics. Figure reproduced from [15]

Is an enterprise solution, guaranteed by the DSI department within Iper company, the advantage of this solution could be found in the fact that all the reports are shared, evaluated, validated and are metrics standardized with the collaboration of the DSI department and the management control department, because from those metrics the analysis to be done are critical in order to take fundamental decision about the company future steps. Being standardize and static reports all the users that needs to do some analysis from the information of Iper products, Sales and so on, need to have access to the MicroStrategy platform even the sales points.



FIGURE 5.14. Microstrategy Level of utilization. Figure reproduced from [15]

The problem of the system that is already known by the users as DWH, is the complexity to get in detail, the system starts to get slowly when the number of references increases, the analysis, and the data discovery are not possible with this system, neither immediate, and the majority of users express the interest in having available a system that allows the user to infer some patterns or trends but in a easy way, without having to use another type of tools as complementary as excel with v-lookups.

5.1.4.1 Microstrategy Properties

When selecting the report in the Main Page the system shows a new page, a series of filters (prompts) relevant to the topic of the report. Beyond the label value of the filters, the use of these filters is almost the same. The metrics shown are the results. In some reports they can be chosen, expanding the default or modifying it, or selecting the predetermined and are not even shown in the filters. Product selection filters are provided and start from a choice of the criterion and then present the filter refinement which can be a hierarchy, a manual / file entry of the codes, a list of products. Below there is an example of the choice of the criterion and subsequent insertion of the search values. The tree structure in the report data table allows the "explosion" of the data passing from the XY family to the whole of the XYs that are part of it. at the end is possible to the user identified the interrogation made by Microstrategy to the Iper DWH in order to get the final report and all the selections previously made , in particular there is the report about the daily sales according with the assortment fig(4.5)



FIGURE 5.15. MicroStrategy Prompt details . Figure reproduced from Iper internal documentation.

5.2 Organization Analysis

The current process involves the Data intelligence team inside the digital innovation section in the DSI department which is in charge of providing tools to allow all the members of the company to use them , so, basically has to start with the concept of user friendly , understanding that the people within a company has many kind of backgrounds, studies and ways of thinking so the objective is to achieve all the general requirements allowing them, to increase the autonomy when managing data , here are introduced one of the main final objectives of the team, to implement the concept of self BI inside the compay. But Iper AS-IS situation is a different from the self- BI purposes, when the requirements are presented, there is not a process of evaluation
or approval of these requirements but instead, each member of a given department presents an individual request to the DSI department to develop a solution. The actual mechanism is evidenced in the figure (5.16):



FIGURE 5.16. Team Organization. Figure reproduced from Iper internal documentation.

During the process of business requirements demand there are some flaws identified:

- there are not established formal processes inside each department, for instances in the purchasing department there are 13 buyers each of them, has a certain logic in order to make the analysis about their merchandise structure products in charge, but is not a procedure already shared within the own department.
- All the requests arrive to the DSI department without any kind of previous analysis, there is not evidence of any kind of approval process.
- Re-cycling processes by which meanwhile the request or need is identified, the users doesn't establish the dimensions and metrics by which he wants to effect uate the future analysis.
- The framework is IT center which doesn't allow to increase the agility , this represents the presence of an unique team within Iper that must switch among all the different areas and range business and satisfy all the requirement of the other areas in order to analyze the process.

The unique information about the metrics standardize is the one presented in the reports on Microstrategy and is possible for the user to access to this data in order to understand better and get deeper into each metric concept and then, finally, verify if the metric has been made to do the correct inference about the process. The attributes described are the name and logic of the metrics, the possible selections that can be done and the rules. With all the data collected and visualized through the report sent to the email in excel form from MicroStrategy, the user is able to construct his own analysis according with the metrics needed as it can be seen on the fig (5.17)this analysis is done by the user utilizing tools as excel that makes the analysis procedures complicated and confused.

						-				
SCATOL	AME									
			Week 20 dal 14	/05 - 20/05	vs 15/05	- 21/05	YTD alla Week 20)		
Rep.	Settore	Desc. Gru. Code	VL Week 2018	∆ Cifra Week su 2017	Mrg. Lungo% Week 2018	Dif.Punti Margine Week	VL 2018 YTD	∆ Cifra YTD su 2017	Mrg. Lungo% YTD 2018	Dif.Punti Margine YTD
1	10	10-1 ACQUE E POLVERI								
1	10	10-2 BIBITE								
1	10	10-3 SCIROPPI PER BIBITE								
1	10	10-4 SUCCHI								
1	. 10	10-5 BIRRE								
1	10	10-6 VINOSI								
1	10	10-7 APERITIVI E LIQUORI								
1	10	10-8 RICORRENZE								
1	10	10-9 ACETO, OLIO, ALTRI CONDIMENTI								
1	10 Totale									
1	11	11-0 IGIENE BIMBO								
1	11	11-1 IGIENE CORPO								
1	11	11-2 ASSORBENTI								
1	11	11-3 CURA UOMO								
1	11	11-4 IGIENE CAPELLI								
1	11	11-5 IGIENE ORALE								
1	11	11-6 COSMESI								
1	11	11-7 MAKE UP								
1	11	11-8 DEODORANTI/PROFUMI/RICORRENZE								

FIGURE 5.17. Example Analysis Excel. Figure reproduced from Iper internal documentation.

After all the information is collected and established under the right format to do the analysis, the User starts to construct the presentation to the CIO's in order to give them the right tools to the the decisions. In this case the Users are supported by Excel graphics that in the majority of situations doesn't allow to convey the right message to the end user . For instance, the Management control system elaborates a weekly book and in order to do it, they have just one person in charge working on it the whole week in order to make graphics.

This is a process completed manual so the percentage of risk about human risk is really high and this will convey into wrong insights.

Who is making the analysis?, there are no figures as Data engineers, business analysis or data Analyst, mostly all the people working within the company has started working in Iper for more than 20 years and develop the same activities since a period of 10 or 5 years, so the resistance to change is notorious.

5.2.1 Analytics Maturity

• Data Accessibility: The data is accessible to the ones that need it but through instruments that are not easily to manage and not all of them have access to formation courses in which more than giving the solution also is explained the way to access it. There are not process standardized where is possible to identify the best practices in order to make a given analysis.

- Analytics-Guided Culture: The members of the different department have no culture for the analysis, even in some cases, is not required to make them.
- KPI's established: In some cases there are not KPI's defined neither milestones established. the operational activities are not aligned with the strategy.
- Central Repository: There is a central repository where converges all the data following the *ETL* process, also thanks to the work inside the DSI department there are some implementations in process about the idea of creating a data lake which will operate with the methodology of the *ELT* Process in which all the raw data will be gather and available for further analysis.

DATA STRATEGY Livelli di Maturità TRANSACIONAL MANTICAL Descriptive Automated DATA MONTETIZATION Livelli di Protezione DB Montetization Livelli di Protezione DB

As is described on the figure (5.18) there are different levels of protection of the DB

FIGURE 5.18. Analytics Maturity. Figure reproduced from Iper internal documentation.

5.2.2 Criticalities

According to the situation previously explained, it can be said that Iper has the following criticalities grounded on the requirements for performance measurement metrics, data collection, data governance, Human resources, and analytics processes

- 1. The actual performance metrics are not structured and do not cover all the process.
- 2. There is no correlation metrics based on strategy and/or objectives established by the company.
- 3. The metrics do not relate decision levels, strategic, tactical and operational.
- 4. Not all the metrics established by the individual employees allows to follow a given target according with Iper strategy.

- 5. Data collection and calculation methods are not clearly detailed, dismissing the data governance.
- 6. There are not enterprise tools given by the DSI department available to all Iper that allow users to make easy and immediate the analyst daily work.
- 7. There are not figures of Business Analyst inside each department.
- 8. There is not a system to generate agility inside the DSI department in order to collect and answer to all the colleagues requests.

The critical problems presented are justified by the defects in the mapping of the process, the management objectives, the structure of the information system and the availability of data. For instance, there was no process mapping of the activities done in each of the departments, additionally the information system structure for the users is not updated and does not provide enough infrastructure in order to make use of the tools given by the DSI department .All together are some of the causes behind the critical problems.Therefore, the understanding of the AS-IS situation and the criticisms detected at the time, provide a starting point for the selection of the best BI tool that will be adapted and that meets the requirements of the users.

Chapter 6

To Be: Qliksense

During this chapter it is presented the Qliksense tool which opens up the discussion of the **TO-BE situation** of Iper Montebello Data Analysis processes, their implementation and management (later analyzed in detail inside the Implementation chapter (7). There are presented the main characteristics of the Qliksense system, there is shown an overview of Qlik as software product (6.1).

6.1 Overview

Qliksense is a BI tool that allows the user to answer questions that weren't thought in advance, is an specialized system that gives all the tools to the user so it is possible to effectuate data discovery and to make decisions thought the use of the visualizations based on selections. Qliksense works creating different applications by which the visualizations will correspond to a given collection of data (Dimensions, measures), in this way it is possible to gather the data in a self-contained file ,with a certain structure, QVF format file with the purpose of containing the information from the business, necessary and useful to make the consecutive analysis. The applications are connected to a data source which is the provider to get the different information from the business facts; the information is reusable and it is possible for the user to develop some expressions in order to effectuate some calculations , create or name in a customized format the metrics (the one that are needed to measure) and the dimensions by which this data is going to be measured.

Apart from the possible interactions accessed by the user, the system, according to the model of enterprise deployment elected by the client, can be utilized by different users at the same time with an optimal performance, the election will depend on the perimeter of use and the future possible evolution to be done inside a company, the differences are relative to the number of servers by which the system is configured to run

	Single-node (small)	Multi-node (medium)	Multi-node (large)
Apps	50	100	1000
Active apps per day	25	50	125
Total users (from UDC)	500	1000	50000
Concurrent users (equals active users within the same hour)	50	100	500
Average app size (in gigabytes)	0.1	0.1	0.1
Maximum app size (in gigabytes)	1	2	5
Content creation (objects per hour)	20	40	50
Reloads per hour	10	20	400

FIGURE 6.1. Enterprise Deployment. Figure reproduced from [26]

One of the main advantages of Qliksense, is the capability to read the data from different type of formats from a Database table or view to Html tables, but although the applications read the data from this kind of sources the configuration of the system permits the data architect to define the relationships instead of the common method of establishing (common fields- keys), those common fields are automatically defined taking in to account the name and there is not previous data aggregated.

To the user is offered a wide range of possibilities that are there to be discovered through data visualizations whenever it is considered necessary by the user, it is possible to make operations of pivoting, gaining instantaneously insights without the help of a technician., the system will allow to evaluate a huge spectrum of cases.

Eliminating the communication silos is one of the main purposes of this solution, the system allows the users to share data in real time, being connected by a unified and safe hub. This is traduced by a uncountless number of possibilities. When analyzing data from a huge company with daily records it is fundamental to have a tool by which should be possible to analyze massive numbers.

The possibilities with this product are several this is one of the main reasons by which the options offered out stands comparing the market competitors, the basic features, the continuing innovative solutions and evolutions make from it one of the best alternatives related to the BI solutions for the retail trade companies.

6.2 Get close to the application

There is a desktop application developed by Qlik, it is free and is an useful instrument in order to get close with the user, it will allow to use with confidence the different functionalities of the solution , it is easy to use, therefore it allows also analyst to access to the different services in the same way an enterprise solution allows. The Enterprise version enables the possibility to edit all properties that are necessary for the technical execution, to modulate the data, and effectuate the complete process.

6.2.1 Main Characteristics

Qliksense in order to provide the user of a certain environment where is possible to navigate through all the applications created and become familiar with it has created the hub,



FIGURE 6.2. Qliksense Hub. Figure elaborated with Qliksense Desktop

The hub has the structure that is being shown in the Figure (6.2), the environment is visually friendly and allows the user to create also his own applications. For creating and application it is possible to do it by to different ways:

 Drag and drop data load: This is a functionality that will be used by the final user interested about making some quickly inferences and it doesn't required any precedent skill in order to do it; it is about load local data into the app by dragging and dropping files and selecting the information from these files. 2. Data load editor: is the most powerful tool that will allow the analyst to set up the data registered in the application, to format the data in a specific way, even to create some metrics that are fundamental for the core business and that will be utilize with a certain frequency. It is in the Data load editor where the connection to the information source is made, and the script can be generated automatically or type in a manual form.



FIGURE 6.3. Qliksense Create an Application. Figure elaborated with Qliksense Desktop

After Loading the data it is possible to access to :

Application View



In this section is possible to have access to the different pages created within the application, to have an overview about the title, description and last time by which the data was loaded.

Application View Figure elaborated with Qliksense Modeler[26]

Data Manager



Data Manager Figure elaborated with Qliksense Modeler[26] By selection from the menu this option the user will have the visibility of the information that the application contain, The possible associations, the number of tables available, the type of load editor function utilized when loading it, and even the data itself, the fields and information within them.

Data Load Editor



Data Load Editor Figure elaborated with Qliksense Modeler[26]

After the connection is established it is possible to extract the data through an SQL select statement, furthermore the script is also user

friendly because with the colors allows the analyst to identify the different components of the programming syntax, following the best practices the user can name the tables according with the specific requirements, versioning format is allowed, comment the code is possible and also rename the fields. As the application works based on the sql language it is possible to make the common operations able as Inner, Outer, Left and right join, concatenations, aggregations and so on.

Qliksense provides a debugger environment where the user is able to identify and isolated errors by selecting the outputs, variables or eventually breakpoints that can be inserted or added by the analyst according with the circumstances.

Data Model Viewer

	Divisions				Categories	
	P DivisionID >				CategoryID 🔎	h
	DivisionName				CategoryName	
Customers					Description	Products
DivisionID 🔎	4	Orders		l		CategoryID #
CustomerID 🔎		CustomerID 🔎	11	OrderDetails		ProductID #
Address		Order10 🔎	h	ProductID 🌶	-	ItemCost
City		EmployeeID	4	Order1D 🔎		ItemPrice
CompanyName		FreightWeight	11	Discount		ProductName
ContactName		OrderDate		LineNo		QuantityPerUnit
Country		Year		Quantity		Supplier1D
Fax		Month		UnitPrice		UnitCost
Phone		Day	11	LineSalesAmount		CataloguePrice
PostalCode		ShipperID				UnitsInStock
StateProvince						UnitsOnOrder

Data Model Viewer Figure elaborated with Qliksense Modeler[26] In this Section, it is possible to navigate between the tables, to verify the keys created automatically, to get in detail about the fields filled, the density indicator, the corresponding relations between tables, it is possible to compare data structures, identify the present distinct values and number of rows and columns populated.

It is possible to make use of some options within the data model viewer as the collapse or expand options.

Iper Montebello company has the possibility of implementing activities of analysis by means of this tool, there are many advantages but the most common are: letting the users to experiment and get confidence with the instrument. Users can be anyone inside business areas. The application allows the user to customize the appearance and also to make use of visualizations already created and standardize so the aspects that are relevant, depending on the requirements of each business area, can be more easily accessed and therefore it could be possible to generate a new process that should add value to the company, at the same time the reports and the presentations about the facts can be optimized.

Chapter 7

Methodology and Implementation Plan.

This chapter contains all the data related to the implementation process achieved in the Iper Montebello company for the business analysis process and the collection of requirements, regarding the activities of the different departments among the company. First of all, it is mentioned the As-Is situation and the criticalities found on it, this means that it is shown a description of each of the actual tools used by the users in order to take decisions and make some confronts with other areas from the business there are also established some inferences to highlight within the scope of this dissertation, all of these in order to improve and optimize the enterprise actual processes and to defined new ones, At the end it is presented the To-Be situation, selected considering the literature proposed on the chapters (2 and 6), The comparison is done , the implementation process on Qliksense, and a description on how the organizational structured will be changed.

7.1 AS-IS Partial conclusion

As previously discussed, in order to relate business and Techniques, to reduce re processing, confusion, to increase co-working activities inside Iper, to growth in emerging markets and to ask the right questions is necessary to implement a new tool always with the goal of evolving, the solution after the previous analysis of the competitors and current offer in the market is called Qliksense and as has been described in the previous chapter (6), the solution was acquired by the company during the month of June 2018, configured and installed by the infrastructure team inside the DSI departmen. This tool will allow to satisfy all the requirements from the different users that were identified during the As-Is chapter, so the process of data extrapolation will be immediate and simple. Processes that from 10 years ago have been done through the use of not enterprise instruments as:

- CercaCodice: as described during the subsection (5.1.1).
- Webgate as described during the subsection (5.1.2).
- Twittiper as described during the subsection (5.1.3).

This tool is mainly proposed under the basis of the data discovery functionality, that couldn't be achieved before during the use of the current user tools inside Iper; the idea is to cover the whole perimeter of the activities done with the applications and to improve the users experience. all the planning framework was made doing user center logic, because in particular the BI tools, have to correspond with the users expectations, this will be automatically transferred as an internalization process and a positive shift of the learning curve.

During the planning step in specific is really important to focus on the necessity of having a multidisciplinary team because in most of the cases the solutions made by the technicians have an adoption percentage really small, here the importance reside in the final user involving process during all the steps of the project. On the other hand there is an organizational and structure argument that will be covered further on, and is the fact of creating new cross-functional process that involves all the agents affected and interested about the tools used to make the analysis techniques.

One of the main purposes of this implementation involves the problem of excess of requirements that arrives every day not only to the data intelligence team within the DSI department but also to other members, where the escalation levels are enough to create bugs and decrease the efficiency of the requirements satisfaction. As a consequences of these behavior the Data intelligence Team has been involved in a reduction process of the number of reports available for the business users (almost 3000) til February 2018, and that after a hard coordination job was able to be reduce to a number of 500 reports available on the MicroStrategy system, as has been explained during the subsection (5.1.4) the only one certificated by the DSI department but that unfortunately just allows to have access to some statics reports. All of the problems exposed regard not just the design of each process but also to the lack of tools available and the lack of coordination between the interested parties. Apart of the deficiencies and disadvantages found on the lack of instruments, the fact of having different tools that aren't enterprise resources, that doesn't ensure the quality and veracity of the data and avoiding the sharing process between the departments; This makes clear the necessity of Iper about adopting mechanisms and instruments that provide competitive advantages, in terms of availability, efficiency and performance; also from a perspective of cost that will be reduce when implementing a unique solution that doesn't involves the use of Virtual machines and at the same time increase the economies of scale. the idea comes also from a study made by the data intelligence team but also above all from all the request made by the employees inside the company.

7.2 Implementation

The overall project plan for the BI implementation can be identified in the figure (7.1)



FIGURE 7.1. Overall Project Plan.

The focus of this work correspond to analyze the solution design til the Kickoff, the data mapping process and the definition of the request and implementation process including analyzing the organizational structure.

First of all, it is important to remember that as previously explained in the introduction of this dissertation, the general goal of this implementation is to evolve and to allow the correct development of the different business processes. In this way, Iper will be an organization that can run in a more efficient workflow between the different areas inside the company. Secondly, to understand that the analysis processes are the fundamental differentiators, and this is important mostly when talking about the decisions making process that took place everyday in Iper company. The To-Be situation considers the Qliksense system which will satisfy all the current requirements and task that should perform the business agents and in contrast with the current solutions will allow to insert within the company a new analysis and data extracting tool. To increase the agility on the different process by which the user is involved, where her/him will be able to:

1. The user will have the possibility to create different visualizations in an autonomy way

what it is really difficult to achieve with the solutions presented, the system is designed thinking in the final user.

- 2. The users are able to navigate trough the visualizations making use of tables, graphs and so on , the perimeter of excel and the previous tools will be covered with functionalities as pivot tables that will permit to have access to the data in real time.
- 3. The user will be able to select the metrics and dimensions by which will want to center the process of data discovery. To illustrate, the most common requirement from the users is the visualization of the sales and the stock at the same time. with the visualizations that already exist they will just need to be mapped or modified by the user (Business Analyst) in this case will be capable to create the metrics on the front end (directly with the application functionalities).
- 4. It is possible for the user to export the data in different formats as excel in the case of massive data load, also in the case of needing to prepare a presentation it is possible to do it within the application it self and to download it, the pdf files are also made directly in it.

Microstrategy won't be substituted by Qliksense because it will cover other type of user request.

7.2.1 Qliksense Implementation

Constraints for back-end implementation:

- There are some constraints about how the information in the DWH was set, because some of the tables moreover that storage useful information also gathers data that is not worth to measure for the process that is been analyzed and because of this is not possible to avoid the loading on the applications. This phenomenon will represent some conflicts with the user because will have access to all of it and will just generate confusion.
- There is some information that has not been mapped so it is important to start with the information fluids to be transferred to the governance DWH system.
- There are some analysis for the purchasing department where the information is not trace, to illustrate the price of a product in the DWH is visualize showing just the last price but the chronology is not presented on the dates.

The BI solution works based on QVF files where the applications are created, but in order to optimized the loading of the data it was necessary the creation of QVD files where just for once the data has to be loaded and then when an application that needs this information is created, the user just have to load the QVD file and the time will be Faster up to 10-100 times than other sources. Due to the fact that the algorithms use to store data in memory. In order to create these QVD files it was necessary to map all the information useful for the different analysis of the

purchasing department and not just for the single application. The files are divided by Facts (Sales, purchasings, movements and so on) and Registry (Product, Sales points, Provider and Calendar) as follows.

- 1. Registry Tables:
 - Anagrafica Prodotto: Is the table in which is defined all the information required in order to identified a product and it's insert process into the different systems, in order to optimized the loading time it was necessary to report also the information from other tables thought the keys defined on the dwh **'ID'** it can be visualized in the Appendix (A.1).
 - Anagrafica Fornitore (Registry of the providers): In this table created in the QVD files was reported all the information of the providers from the key ID to the brand by which is classified the Provider. To get in detail refer to the Appendix (A.2).
 - Anagrafica Ente: is the table in which is defined all the information required in order to identified a sales point and it's insert process into the different systems, in order to optimized the loading time it was necessary to report also the information from other tables thought the keys defined on the DWH as the area description from the table Ana.Aree, the file was saved under the path of the user that created it and then will be used in order to report the information needed in the correspondent applications.To get in detail refer to the appendix (A.3).
 - Calendar: Is the Table in which is reported all the information regarding to the calendar and by which all the facts are registered depending on the granularity of the events this can be aggregated as the correspondent Appendix (A.4).
- 2. Facts Tables:
 - Assortment Information: In this table is registered the information about the stock presented in each of the 27 sales point, with the prices and so on, Appendix (A.5).
 - VendutoPromoGiorno: During this Table where can be identified the attributes regarding to the Sales that where done thorough a mechanism of push demand in which the promotion campaigns where presented, Appendix (A.6).
 - Merchandise Structure: In this table is defined all the merchandise structure with the codes in order to identified them, the relation between the structure dependency and the description of each of them, Appendix (A.7).
 - Stock: In this table has been explicit all the information regarding to the stock sales, the number of pieces presented in each of the stores, and the inventory differences founded, Appendix (A.8).

Each of these QVD where created at first through a QVF file in which the data was loaded and the tables within the QVF where dropped in order to optimize the memory space. Each of the QVD depending on the information contained in it occupy a different amount of memory to illustrate the Fact about the purchasings has close to 17.086.886 lines, the QVD where saved in a certain path from the machine where the BI tool was downloaded and the memory occupied correspond to the following:

Name	Date modified	Туре	Size
Qvd_Acquisti.qvd	8/29/2018 9:15 AM	QVD File	287,669 KE
Qvd_AnagEnte.QVD	8/31/2018 3:52 PM	QVD File	29 KE
Qvd_AnagForn.QVD	8/31/2018 3:52 PM	QVD File	4,008 KE
Qvd_AnagProd.QVD	8/31/2018 3:52 PM	QVD File	290,986 KE
Qvd_AssortimentoPdv.qvd	8/31/2018 5:02 PM	QVD File	132,792 KE
Qvd_AssortimentoPromo.	QVD 8/29/2018 9:13 AM	QVD File	125,227 KE
Qvd_Periodi.qvd	8/31/2018 4:15 PM	QVD File	48 KE
Qvd_StockContabile.qvd	8/29/2018 12:29 PM	QVD File	707,808 KE
Qvd_StockFisico.qvd	9/1/2018 2:30 AM	QVD File	11,939,483 KE
Qvd_VendutoGiorno.QVD	8/31/2018 7:33 PM	QVD File	2,425,614 KF
Qvd_VendutoPromoGiorno	OQVD 8/31/2018 12:10 PM	QVD File	715,686 KE

FIGURE 7.2. Memory occupied by the QVD files.

Right after the data within the DWH was mapped and saved in the memory it is necessary to construct the Application having into account the customer requirements and expectations here has to be explicit that not all the information mapped will be used, but in another case it will be possible to find it and the business analyst will be able to use it according with the necessity. There are different Subsections defined during the application creation following the best practices:

• Main: At first have to be defined the format by which the BI tool is gonna read the data, also the variables of the application and the connections established in order to capt the information from the DWH:

```
SET ThousandSep=',';
SET DecimalSep='.';
SET MoneyThousandSep=',';
SET MoneyFormat='$#,##0.00;-$#,##0.00';
SET TimeFormat='h:mm:ss TT';
SET DateFormat='M/D/YYYY';
SET TimestampFormat='M/D/YYYY h:mm:ss[.fff] TT';
SET FirstWeekDay=6;
```

```
SET BrokenWeeks=1;
SET ReferenceDay=0;
SET FirstMonthOfYear=1;
SET CollationLocale='en-US';
SET CreateSearchIndexOnReload=1;
SET MonthNames='Jan;Feb;Mar;Apr;May;Jun;Jul;Aug;Sep;Oct;
Nov;Dec';
SET LongMonthNames='January;February;
March;April;May;June;July;August;
September;October;November;December';
SET DayNames='Mon;Tue;Wed;Thu;Fri;
Sat;Sun';
SET LongDayNames='Monday;Tuesday;
Wednesday;Thursday;Friday;Saturday;Sunday';
//Connections
LIB CONNECT TO 'DWH (sense_digital_innovation1)';
// MyVariables
SET vStartDate = 20170101;
```

```
SET vEndDateRicalcolo= 20180531;
SET vStartDateDiffe= 20180601;
SET vEndDateDiffe = 20180617;
```

• After declaring the Variables it is necessary to call the information (Facts) by which are gonna be made the corresponding analysis and the measurements are based on also was necessary to make some transformation to the information and to create different tables in order to access to it because it's not ready to be used for the user in the way that is presented in the DWH:

```
// 1. Outer Join between LinkTableNC - NoteCredito(FACT_ACQ) - Sales
// 2. The Resident option was used in order to read the table created.
trace 'Creazione Vend_microsett_ric_prod_int....';
```

//Loading Note di Credito
[VendutoNoteCredito]:

```
LOAD MESE_COMP_ID,
CAUSALE_CD,
PROD_ID,
ENTE_ID,
    ACQ_VAL as NotaCreditoOrig;
SELECT "MESE_COMP_ID",
"CAUSALE_CD",
"PROD_ID",
"ENTE_ID",
    "ACQ_VAL"
FROM "FIN_DATA"."FACT_ACQ" acq WHERE CAUSALE_CD='09' and
MESE_COMP_ID>=$(vStartDate);
join(VendutoNoteCredito)
//Loading link note di credito - Sales
[NoteCredito_join]:
LOAD * Inline [
MESE_COMP_ID, MICROSETTIMANA_ID, NC_KEY, MICROSETTIMANA_ID_NC
20170101, 2017061, 201701012017061, 2017061
20170201, 2017101, 201702012017101, 2017101
20170301, 2017141, 201703032017141, 2017141
20170401, 2017181, 201704012017181, 2017181
20170501, 2017231, 201705012017231, 2017231
20170601, 2017271, 201706012017271, 2017271
20170701, 2017321, 201707012017321, 2017321
20170801, 2017361, 201708012017361, 2017361
20170901, 2017401, 201709012017401, 2017401
20171001, 2017451, 201710012017451, 2017451
20171101, 2017491, 201711012017491, 2017491
20171201, 2017521, 201712012017521, 2017521
20180101, 2018061, 201801012018061, 2018061
20180201, 2018101, 201802012018101, 2018101
20180301, 2018141, 201803032018141, 2018141
20180401, 2018181, 201804012018181, 2018181
20180501, 2018231, 201805012018231, 2018231
20180601, 2018271, 201806012018271, 2018271
20180701, 2018321, 201807012018321, 2018321
```

```
20180801, 2018361, 201808012018361, 2018361
20180901, 2018401, 201809012018401, 2018401
20181001, 2018451, 201810012018451, 2018451
20181101, 2018491, 201811012018491, 2018491
20181201, 2018521, 201812012018521, 2018521
];
outer join(VendutoNoteCredito)
//Loading data of Ricalcolo sectors selected
Vend_microsett_ric_prod_int:
LOAD
    MICROSETTIMANA_ID,
  ENTE_ID,
PROD_ID,
VEN_QTA_PS as VendutoQtaPs,
VEN_QTA_PZ as VendutoQtaPz,
VEN_VAL as VendutoLordo,
VEN_NET_VAL as VendutoNetto,
COSTO_VAL as Costo,
VEN_UTILE as Margine1,
FORN_ID_ACQ_PRINC as FORN_ID,
UTILE_PFA_ACTUAL As MargineLungoNew,
    VEN_UTILE_NET_PFA_NEW as MargineLungoOld,
    UTILE_PFA_ACTUAL As MargineLungo,
COSTO_PFA_ACTUAL as CostoNettatoPfaNew,
    COSTO_VAL_NET_PFA_NEW as CostoNettatoPfaOld,
    COSTO_PFA_ACTUAL as CostoNettatoPfa,
    date(date#(DATA_ELAB,'YYYYMMDD'),'DD/MM/YYYY') as DataElab,
    FORN_TYPE,
    'Ricalcolo' as CALC_TYPE
;
SELECT
periodi.MICROSETTIMANA_ID,
```

```
GI IOUI . MI
```

```
periodi.MESE_ID,
```

```
ven.ENTE_ID,
```

```
ven.PROD_ID,
sum(ven.VEN_QTA_PS) as VEN_QTA_PS,
sum(ven.VEN_QTA_PZ) as VEN_QTA_PZ,
sum(ven.VEN_VAL) as VEN_VAL,
sum(ven.VEN_NET_VAL) as VEN_NET_VAL,
sum(ven.COSTO_VAL) as COSTO_VAL,
sum(ven.VEN_UTILE) as VEN_UTILE,
ven.FORN_ID_ACQ_PRINC,
sum(ven.COSTO_VAL_NET_PFA_NEW) as COSTO_VAL_NET_PFA_NEW,
sum(ven.VEN_UTILE_NET_PFA_NEW) as VEN_UTILE_NET_PFA_NEW,
sum(ven.UTILE_PFA_ACTUAL) as UTILE_PFA_ACTUAL,
sum(ven.COSTO_PFA_ACTUAL) as COSTO_PFA_ACTUAL,
    ven.DATA_ELAB,
    ven.FORN_TYPE
FROM FIN_DATA.PFA_ACTUAL_GIORNO_PROD ven
join DIM_PERIODI periodi on ven.PERIOD_ID=periodi.PERIOD_ID
where ven.PERIOD_ID between $(vStartDate) and $(vEndDateRicalcolo)
group by periodi.MICROSETTIMANA_ID, periodi.MESE_ID, ven.ENTE_ID,
ven.PROD_ID,ven.FORN_ID_ACQ_PRINC,ven.DATA_ELAB,ven.FORN_TYPE
;
Concatenate(VendutoNoteCredito)
//Loading data DIFFE by week just for the sectors with the recalcultaion.
LOAD ENTE_ID,
MICROSETTIMANA_ID,
PROD_ID,
VEN_QTA_PS as VendutoQtaPs,
VEN_QTA_PZ as VendutoQtaPz,
VEN_VAL as VendutoLordo,
VEN_UTILE as Margine1,
VEN_NET_VAL as VendutoNetto,
COSTO_VAL as Costo,
COSTO_VAL_NET_PFA_NEW as CostoNettatoPfaOld,
    COSTO_VAL_NET_PFA_NEW as CostoNettatoPfa,
VEN_UTILE_NET_PFA_NEW as MargineLungoOld,
    VEN_UTILE_NET_PFA_NEW as MargineLungo,
FORN_ID_PRINC_ACQ as FORN_ID,
```

```
'FORN_DIFFE' as FORN_TYPE,
    'Diffe' as CALC_TYPE;
SELECT
periodi.MICROSETTIMANA_ID,
    periodi.MESE_ID,
ven.ENTE_ID,
ven.PROD_ID,
sum(ven.VEN_QTA_PS) as VEN_QTA_PS,
sum(ven.VEN_QTA_PZ) as VEN_QTA_PZ,
sum(ven.VEN_VAL) as VEN_VAL,
sum(ven.VEN_NET_VAL) as VEN_NET_VAL,
sum(ven.COSTO_VAL) as COSTO_VAL,
sum(ven.VEN_UTILE) as VEN_UTILE,
ven.FORN_ID_PRINC_ACQ,
sum(ven.COSTO_VAL_NET_PFA_NEW) as COSTO_VAL_NET_PFA_NEW,
sum(ven.VEN_UTILE_NET_PFA_NEW) as VEN_UTILE_NET_PFA_NEW
FROM FIN_DATA.FACT_VEN_GIORNO_PRODOTTO ven
join DIM_PERIODI periodi on ven.PERIOD_ID=periodi.PERIOD_ID
join DIM_PROD prod on ven.PROD_ID=prod.PROD_ID
where ven.PERIOD_ID between $(vStartDateDiffe) and $(vEndDateDiffe)
and prod.SET_CD in
('50',
'71',
'11'.
'12',
<sup>'81'</sup>.
'66',
'16'.
'20',
'31',
<sup>'33'</sup>,
'10',
'51',
'56',
'32',
'14',
```

```
'15',
'37',
'52<sup>'</sup>,
<sup>'80</sup>,
'40',
'64',
'24',
<sup>'53'</sup>.
'30')
group by periodi.MICROSETTIMANA_ID,periodi.MESE_ID,ven.ENTE_ID,
ven.PROD_ID,ven.FORN_ID_PRINC_ACQ
;
Concatenate(VendutoNoteCredito)
//Loading data for the sectors that don't have the recalculation.
LOAD ENTE_ID,
MICROSETTIMANA_ID,
PROD_ID,
VEN_QTA_PS as VendutoQtaPs,
VEN_QTA_PZ as VendutoQtaPz,
VEN_VAL as VendutoLordo,
VEN_UTILE as Margine1,
VEN_NET_VAL as VendutoNetto,
COSTO_VAL as Costo,
COSTO_VAL_NET_PFA_NEW as CostoNettatoPfaOld,
    COSTO_VAL_NET_PFA_NEW as CostoNettatoPfa,
VEN_UTILE_NET_PFA_NEW as MargineLungoOld,
    VEN_UTILE_NET_PFA_NEW as MargineLungo,
FORN_ID_PRINC_ACQ as FORN_ID,
    'FORN_DIFFE' as FORN_TYPE,
    'Diffe' as CALC_TYPE;
SELECT
periodi.MICROSETTIMANA_ID,
    periodi.MESE_ID,
ven.ENTE_ID,
```

ven.PROD_ID,

sum(ven.VEN_QTA_PS) as VEN_QTA_PS, sum(ven.VEN_QTA_PZ) as VEN_QTA_PZ, sum(ven.VEN_VAL) as VEN_VAL, sum(ven.VEN_NET_VAL) as VEN_NET_VAL, sum(ven.COSTO_VAL) as COSTO_VAL, sum(ven.VEN_UTILE) as VEN_UTILE, ven.FORN_ID_PRINC_ACQ, sum(ven.COSTO_VAL_NET_PFA_NEW) as COSTO_VAL_NET_PFA_NEW, sum(ven.VEN_UTILE_NET_PFA_NEW) as VEN_UTILE_NET_PFA_NEW FROM FIN_DATA.FACT_VEN_GIORNO_PRODOTTO ven join DIM_PERIODI periodi on ven.PERIOD_ID=periodi.PERIOD_ID join DIM_PROD prod on ven.PROD_ID=prod.PROD_ID where ven.PERIOD_ID between \$(vStartDate) and \$(vEndDateDiffe) and prod.SET_CD not in ('50', ⁷¹ '11', '12', '81', *'66'*, '16', '20', '31', ³³ '10', '51', ^{'56'}, '32'. '14', '15'. '37', '52', '80', '40', '64', '24', '53', '30')

group by periodi.MICROSETTIMANA_ID,periodi.MESE_ID,ven.ENTE_ID,

```
ven.PROD_ID,ven.FORN_ID_PRINC_ACQ
;
[Vend_microsett_ric_nc_prod]:
LOAD
    MICROSETTIMANA_ID,
    MICROSETTIMANA_ID_NC,
    MESE_COMP_ID,
  ENTE_ID,
PROD_ID,
VendutoQtaPs,
VendutoQtaPz,
VendutoLordo,
VendutoNetto,
Costo,
Margine1,
FORN_ID,
CostoNettatoPfaOld,
MargineLungoOld,
MargineLungoNew,
    MargineLungo,
CostoNettatoPfaNew,
    CostoNettatoPfa,
    DataElab.
    FORN_TYPE,
    CALC_TYPE,
    (if(IsNUll(NotaCreditoOrig), 0, NotaCreditoOrig))as NotaCredito,
    (MargineLungo - (if(IsNUll(NotaCreditoOrig), 0, NotaCreditoOrig)))
    as MargineLungoNC
    Resident VendutoNoteCredito;
```

• Anagrafica (Registry): Here we call the QVD previously created in order to optimize the loading process, in special of the data registered that is needed to identify the facts, are

```
know as dimensions as follows:
    [DIM_PROD]:
LOAD
    PROD_ID,
    PROD_CD,
PROD_DESC,
RET_UNIT_CD,
SUBFAM_CD as Sottofamiglia,
FAM_CD as Famiglia,
GRUP_CD as Gruppo,
QTA_FLG,
IVA_VAL,
REP_CD as Reparto,
SET_CD as Settore,
BRAND_FORN_CD as Brand,
MARCHIO_TIPO_CD,
MARCHIO_CD,
    TAC_CD as TIPO_ASSORTIMENTO_CD,
    PANIERE_CD as Paniere
    where Exists(PROD_ID);
SELECT
    "PROD_ID",
    "PROD_CD",
"PROD_DESC",
"RET_UNIT_CD",
"SUBFAM_CD",
"FAM_CD",
"GRUP_CD",
"QTA_FLG",
"IVA_VAL",
"REP_CD",
"SET_CD",
"BRAND_FORN_CD",
"MARCHIO_TIPO_CD",
"MARCHIO_CD",
    "TAC_CD",
    "PANIERE_CD"
```

```
FROM "FIN_DATA"."DIM_PROD";
```

CHAPTER 7. METHODOLOGY AND IMPLEMENTATION PLAN.

[ANA_PROD_MARCHIO]: LOAD MARCHIO_CD, MARCHIO_DESC where Exists(MARCHIO_CD); SELECT "MARCHIO_CD", "MARCHIO_DESC" FROM "FIN_DATA"."ANA_PROD_MARCHIO";

[ANA_PROD_MARCHIO_TIPO]: LOAD MARCHIO_TIPO_CD, MARCHIO_TIPO_DESC where Exists(MARCHIO_TIPO_CD); SELECT "MARCHIO_TIPO_CD", "MARCHIO_TIPO_DESC" FROM "FIN_DATA"."ANA_PROD_MARCHIO_TIPO";

[ANA_TIPO_ASSORTIMENTO]: LOAD TIPO_ASSORTIMENTO_CD, DESCRIZIONE as TIPO_ASSORTIMENTO_DESC; SELECT "TIPO_ASSORTIMENTO_CD", "DESCRIZIONE" FROM "FIN_DATA"."ANA_TIPO_ASSORTIMENTO" WHERE FLAG_IU!='D';

```
[ANA_PROD_COD_PANIERE]:
LOAD PANIERE_CD,
PANIERE_DESC;
SELECT "PANIERE_CD",
"PANIERE_DESC"
FROM "FIN_DATA"."ANA_PROD_COD_PANIERE";
```

[DIM_FORN]: LOAD FORN_ID,

```
FORN_CD,
FORN_DESC,
FOAM_CD,
FOAM_DESC
    where Exists(FORN_ID);
SELECT "FORN_ID",
"FORN_CD",
"FORN_DESC",
"FOAM_CD",
"FOAM_DESC"
FROM "FIN_DATA"."DIM_FORN";
[DIM_ENTI]:
LOAD ENTE_ID,
ENTE_CD,
ENTE_DESC
    where Exists(ENTE_ID);
SELECT "ENTE_ID",
"ENTE_CD",
"ENTE_DESC"
FROM "FIN_DATA"."DIM_ENTI";
[DIM_PERIODI]:
LOAD
MICROSETTIMANA_ID,
Left(MICROSETTIMANA_ID,4) as AnnoVendita,
Num(Mid(MICROSETTIMANA_ID,5,2)) as SettimanaVendita,
Right(MICROSETTIMANA_ID,3) as MicrosettimanaVendita,
Num(Right(MESE_ID,2)) as MeseVendita;
SQL SELECT
"MICROSETTIMANA_ID" ,
MESE_ID
FROM "FIN_DATA"."DIM_PERIODI"
where PERIOD_ID>=$(vStartDate) and PERIOD_ID<=$(vEndDateDiffe);</pre>
```

• Drop section: These section was created under the purpose of cleaning and eliminating the

information from the memory that was necessary in a certain moment to create the metrics but that for the aim of the applications are not more needed.

drop table VendutoNoteCredito;

7.2.1.1 Qliksense Solution

Qliksense is accessible through web under the appropriate security procedures applied in order to avoid external people to connect and to take advantage from the information, this doesn't affect the possibility of establishing connection making use of different devices as telephones, tablets and so on. When starting the interaction for the final user there will be available different types of streaming

Inside of each of the folders in the main stream it is possible to have visibility of the applications according with the different accounts created and users of the system:

When opening the application:

The application will contain the dimensions and the metrics about the information needed for the users. When opening the application it is possible to identify the name, the description of the information contained on it, the last date in which has been made the updates of the data load. Qliksense is a tool that reads the data from the DWH and gathers it on the memory, it is important to declare that as part of a whole new process was necessary to establish the procedures and the scheduled for the data load but obviously having into account when the information is available in the dwh after the enterprise process had been completed. *When navigating through the application*:

The Iper users have the possibility of navigating over the different visualizations, there are two types described as follows:

• Base pages: There are available some visualizations created by the Data Intelligence team with all the metrics and dimensions already established and are visible according with the user accounts configuration. For the purposes of the project and for a greater trace ability it was created an introduction visualization page, by which the current user can identify which dates the information is related to, the changelog panel where are exposed the changes from the first version of the application, those changes are considered as evolutions because from the data of release the application will suffer some modifications according with the user acceptance tests and the future requirements. So the improvements are transparent. For the purpose of the first phase delivery, was necessary to create an application about the Iper information related to the registry, the assortment and the sales of the last year, between 2016511 (Data related to the micro week, so week number 51)and 20183501, the scope is to identify the information by period, sales point and product.

Introduzione

Margine Lungo Ricalcolo e DIFFE - v1.0

Margine lungo con dettaglio settimanale unendo i dati ricalcolati mensilmente in modalità manuale coi dati della DIFFE giornaliera. La metrica Margine Lungo è popolata seguendo la seguente priorità:

1) Se presente, il valore del dato col Ricalcolo 2) Dato Diffe stessa logica segue la metrica **Costo Nettato PFA**.

La metrica Margine Lungo NC è calcolata aggiungendo al Margine Lungo a livello microsettimana il valore della Nota di Credito mensile unendo le due granularità (settimana-mese) in base alla tabella in basso "Link Note Credito Mensili - Venduto MicroSettimana"

Le logiche concordate per il ricalcolo sono:

I dati di venduto (lordo e netto), costo, margine lungo old, margine lungo new sono calcolati a livello giornaliero per prodotto-ente
 Il costo del venduto è calcolato in base al costo ultimo immesso proveniente da NSDR

I fornitore principale d'acquisto per ente-prodotto è il fornitore dal quale il pot ha acquisto la referenza maggiormente nell'anno 2018. In caso di acquisto verso il FoCo di piattaforma, si risale al FoCo principale d'acquisto della piattaforma. Nel caso in cui il PDV non abbia acquistato il prodotto nel 2018, si considera il fornitore principale d'acquisto in anagrafica sul DWH (tabella DIM_PROD_PDV) - Le PFA applicate sono 'FA-INC-PERC', 'PROM-BASE-PERC', 'PA-MT-PERC', 'FA-MT-PERC'

E' possibile identificare i dati ricalcolati da quelli della DIFFE utilizzando il campo Modalità Calcolo

Dati Ricalcolati dalla m 201652	icrosettiman 22	a Dati Ricalcolati fino 2018	alla microsettimana 3351	Dati DIFFE fino alla microsettimana 2018361				
Data elaborazione ricalcolo pe	r settore	Link Note Credito Mensili - Vendu	to MicroSettimana	Changelog				
Sattora	Q Elabo	Mese Nota Credito Q	MicroSettimana Venduto Q	Novità versione 1.0				
	Liabu	20170101	2017061	- Ricalcolo su tutti i settori				
-1 - MANCANTE		20170201	2017101	 Aggiunte descrizioni struttura merceologica 				
- DETERG/IGIENE/ACCESS.CASA - ANIMALI (SOLO SU AUTORIZZAZ)		20170301	2017141	Novità versione 0.9				
		20170401	2017181	- Aggiunta metrica Qta Kg/l ref. vend. pz				
4 - SCATOLAME DOLCE		20170501	2017231	 Caricamento dati ricalcolo fino all'ultima 				
11 - PROFUMERIA		20170601	2017271	microsettimana disponibile (fine mese)				
18 - A DISPOSIZIONE(VEDI SETT.37	7)	20170701	2017321	Novità versione 0.8				
15 - SCATOLAME SALATO		20170801	2017361	 Aggiunte metriche in tabella settimana 				
13 - PARAFARMACIA		20170001	2017/01	Novità versione 0.7				
10 - LIQUIDI		20171001	2017451	 - Note di Credito inserite anche su prodotti non venduti e 				
20-		20171001	2017431	su Ente P80				
GASTRON/PESCH/SALUM CONFEZ	ION	20171101	2017491	- Creazione tabelle "Tabella Settimana Margine Lungo"				

FIGURE 7.3. Application introduction page.

Eventually as a request presented by the director from the purchasing department will be integrated an user FAQ page by which the common questions will be answered, so the self learning process is also pushed. All of the base pages count with some filters predefined and according with their selection the information is updated automatically, there have been established some KPI's according with the business needs.

The detail level of the information according with the data reported during the subsection (7.2.1):

- Calendar: it details from the day til the year when the facts have been presented.
- Product: The products are classified under an structure, so it is possible to make analysis about each of the levels of detail finishing with the product itself. In this application called Margine Lungo v1.0 made for one of the departments in order to evaluate the profits with a certain detail level were defined some common filters in each of the pages, so it is possible to effectuate efficient researches. Before developing the application some meetings with the department and the end users (in charge of making the analysis process were needed. The result involved the implementation of two different types of filters:

* Standard Filter (By merchandise structure and by Calendar) as can be appreciate in the figure bellow:



FIGURE 7.4. Application Standard Filters.

- Variable Filters: In order to answer the restriction about the data that is not present in the DWH, in special, the Year to Date. We have followed the logic of the Variables that allows to Search information according to the calendar in special about the week that is where the analysis is done. So the user has to insert the week by wich want to do the analysis, The result can be appreciated as follows Tabella Settimana Margine Lungo

ATTENZIONE filtri periodo

NON utilizzare i filtri su "Anno vendita", "Mese vendita", "Settimana vendita". Utilizzare le variabili "Anno vendita var" e "Settimana vendita var"

Anno vendita var	Sett	imana vendita var	Reparto		Settore	Gruppo	Pro	dotto				
2018	25											
Margine Lungo senza NC												
											Dif.	
Reparto	Q	Settore	Q	Gruppo	Q	Venduto Lordo	∆%VL	Incide % ML	Dif.P ML	Incide % M1	Punti M1	Venduto Lon
Totali												
01 - SCATOLAME		12 - DETERG/IGIENE/ACCES	SS.CASA	4								
01 - SCATOLAME		12 - DETERG/IGIENE/ACCES	SS.CASA									
01 - SCATOLAME		12 - DETERG/IGIENE/ACCES	SS.CASA	1								
01 - SCATOLAME		12 - DETERG/IGIENE/ACCES	SS.CASA									
01 - SCATOLAME		12 - DETERG/IGIENE/ACCES	SS.CASA									
01 - SCATOLAME		12 - DETERG/IGIENE/ACCES	SS.CASA	1								
01 - SCATOLAME		12 - DETERG/IGIENE/ACCES	SS.CASA	1								
01 - SCATOLAME		12 - DETERG/IGIENE/ACCES	SS.CASA									
01 - SCATOLAME		12 - DETERG/IGIENE/ACCES	SS.CASA									
01 - SCATOLAME		12 - DETERG/IGIENE/ACCES	SS.CASA									
01 - SCATOLAME		16 - ANIMALI (SOLO SU AU	TORIZZAZ)									
01-SCATOLAME		16 - ANIMALI (SOLO SU AU	TORIZZAZ)	1								
01 - SCATOLAME		16 - ANIMALI (SOLO SU AU	TORIZZAZ)									
01 - SCATOLAME		16 - ANIMALI (SOLO SU AU	TORIZZAZ)	4								
01 - SCATOLAME		16 - ANIMALI (SOLO SU AU	TORIZZAZ)	1								
01 - SCATOLAME		16 - ANIMALI (SOLO SU AU	TORIZZAZ)	1								
01 COATOLAME		16 ANTMALT/COLO CU AU	TOD17747)									

FIGURE 7.5. Application Variable Filters.

It is also possible to effectuate research with free text and the system automatically allows to find all the possible data related to the input.

• Personalized pages: In order to effectuate this operation it is necessary to log in with a credential of an advance user. It is possible to create pages starting from a blank page or it is possible to duplicate the existing ones and to customize it, adding new information depending on the preferences about the dimensions and metrics that will be drag, in order to create the visualizations on shape of graph text and much more.

For the purposes of the Analysis process we have created Five pages as Follows:

	Argine Lungo Ricalcolo - v8.7 Margine Lungo Ric Uttimi dati caricati: 26 Margine lungo con manuale coi dati de	alcolo - v0.7 giu 2018, 17:14 dettaglio settimanale unen Illa DIFFE giornaliera, integ	do i dati ricalcolati mensilmente rando le Note di Credito.	in modalità				
الله Fogli 🗍 Segnalibri 👌 Racconti								
▼ I miei fogli (5)								
			A A	A 9 8 0 0 0	Crea nuovo			
Introduzione	Barchart con dettaglio	Dettaglio 1	Tabella Settimana Margine Lungo	Tabella Settimana Margine Lungo co	foglio			

FIGURE 7.6. Application. Different pages.

- 1. Introduzione: This page has been already defined in the previous section and the idea is to allow the user identify the information presented in the application.
- 2. Barchart con detaglio: Here it is possible to identify the key metrics to be evaluated as each area requires, the sales during the year by week, the different trends between the years and the margin through the year.
- 3. Detaglio 1 (Detail 1): The Data and metrics more in detail.
- 4. Tabella Settimana Margine lungo : all the metrics needed by the user are shown here by week including the year to date metric.
- 5. Tabella settimana Margine lungo con note di credito: This section of the application is dedicated to show all the relevant metrics including also the credit notes.

The system for the BA allows to visualize all the dimensions that has been mapped and also to create new metrics, making some transformations to the data. The associations are always according to the three main dimensions *product*, *Sales point* and *period*. During the utilization of the application from the final user he is able to identify the dimensions, to create the metrics. To illustrate, The following metric was created in order to meausre the difference in the sold gross between the years:

```
(sum({$<SettimanaVendita={'$(=vSettimanaVar)'},
AnnoVendita={'$(=vAnnoVar)'}>}
VendutoLordo)-sum({$<SettimanaVendita={'$(=vSettimanaVar)'},
AnnoVendita={'$(=(vAnnoVar-1))'}>}VendutoLordo))
/(sum({$<SettimanaVendita={'$(</pre>
```

```
=vSettimanaVar)'},AnnoVendita=
{'$(=(vAnnoVar-1))'}>}VendutoLordo))
```

It is possible to define as standard some metrics that will be used all along the application, as the ones created for the Margin Application by the Data Intelligence team, Also following the best practice the team has defined each metric and specified the purpose and the mechanism of calculation of these metrics so the BA and end user are in the capacity of identifying it:



FIGURE 7.7. Metrics defined for the App.

7.2.2 Organizational Structure

The new implementation considers to involve the efforts of the purchasing team department which are the group of people in charge of trying the first application made by the DSI department team. This involves the creation of an structure that foresees the inclusion of the final users which are the analyst and the CIO's team who at the end will express their needs. Because, currently there is one director from the department and according with the merchandise structure there are 10 different directors which at the same time are in charge of 4 buyers and codify resources each one, that means that it is necessary to unified the process how the requirements are made.

The essence of the problem arrives when it is about organizational culture, quality and data sharing and how the reports and the official information shared with internal and external resources has been managed all this time.

It is pretty difficult when each person creates a personalize solution in order to feed a necessity, without an standardize and certificate data, here the role of the data engineers, analysts and manager of analyst is fundamental in order to create a data governance process and to help the rest of the company to be more data literate.

The problem is evidenced when the company has not structure or processes in order to manage the demand and each department or even worst different people from the same department are not aligned about the driver of the information they need. When something as basics as knowing which is the importance of having the data that are asking for, or the transformation data process involved in order to achieve all the requirements of the business is not clear. The solution process will consist on the nomination of a new committee in charged of validating the requirements done by the directly end users and also the implementation of a new figure role within the company Iper which were not present before with the adoption of the QlikSense Solution, This new role is known as the Business Analyst, as how was described before during the data driven chapter (2), The people under this role will be the bridge between the business functions and DSI department members, they will be the ones that will know the process of both departments and will try to transform the business savvy requirements into the language of the DSI department technicians. Those BA will be inserted in the context of each department and will be constantly in touch with the data intelligence section in order to manage and to take into consideration all the arguments refereed to the reporting and analysis operations. BA skills-focused maturity models emphasize the way companies should re-configure and renew resources to develop new analytic skills, the continual training process allows also to increase these competences, this is a critical process inside the company and should be taken into consideration.

The motivations about taking these decisions are based under the following aims:

- 1. To increase the transparency and clarity about the business requirements for the DSI team.
- 2. To increase the velocity and efficacy of the satisfy requirement operations.
- 3. To increase the autonomy of the business users in order to satisfy the own requirement in terms of reporting and analytics because with the current tools available this possibility is limited.
- 4. To increase the quality of the decisions because of the data quality certification from the information that comes directly from the DWH systems of Iper, so to increase the data governance.
- 5. To rationalize and standardize the number of reports.

The new analytics team will be conformed by:

- Business Analyst: As described before will be also in charge of satisfying the request in autonomy, whenever is the occasion they will be coordinated with the data intelligence team and will have continues upgrading with it.
- The committee: Will define and formalize the requirements in the reporting / Analytics field according to the purchasing department and will request the new developments to the DSI department.
- Team Data intelligence (within the DSI department):
 - 1. Will manage the demand of the developments.
 - 2. Will manage all the activities and development of the Data Base.
 - 3. Will manage all the task related to the BI tool **Qliksense** and MicroStrategy.
 - 4. Will be in charge of the training and collaboration day-to-day with the business Analyst.



FIGURE 7.8. Initial proposal Analytics Team.

It was necessary to standardized the process of requirements, establishing a set of rules that specified how people in Iper are supposed to complete a given task or sequence of tasks. The process description can be identified on the figure (7.9)



FIGURE 7.9. Process proposed to be implemented.

The chosen model should have an structure that fits with the internal processes involving the DataWareHouse architecture and in order to facilitate the Business Analyst job it is crucial to define a set of questions to be answered to make sure the applicant understands the own process and the utility of the measurement as follows:

- Title of the analysis.
- Priority: In order to identify if the request is critical, It's urgent?.
- Petitioner: Who is asking the Report or the change of request.
- Objective: Which is the aim of the petition?.
- KPI: What do you want to analyze?.
- Measure: What do you want to measure?.
- Dimension: The measurement is based on which dimension?.
- Dimension deepness: Which is the deepness level of the measurement.
- Source: From where do we take the data , example: DWH or file FTP.
• Frequency: Which is the frequency of the analysis.

Increase of the culture of analysis of semi structure data are one of the main aims which for sure will be the future in order to localized or get deeper on concepts that maybe in a certain period of time, where non considered relevant. Here is important to evidence, the huge relevance that data is having in these days , not taking in to account the GDPR normative and all the consequences that are carry out about the argument.

Moreover, Obviously there is a change management process included , the change of a culture involves a disruptive process in which is really important to be more business savvy , to have knowledge about the process and the aims of the company because at the end all must be aligne in order to complete with the objectives.

Data driven is not a condition that depends on the time or products, it is a culture that has to grow inside the active resources of the company, is clear that this is a continuity process and the objective is to have an exponential growth, concentrating all the efforts in order to increase people competences and to have a more business analytic company which allows exceptional results to be achieved.

To fully benefit from Information Technology, it is necessary to address the fact that renewing process can be benefit for the company in order to achieve cost savings, improve the operation activities carried out by the employees, specially with all the constant changes.

The re-engineering is an approach implemented among the incumbents, it is based on the idea of managing the processes instead of the functions, the final goal is to improve in an incremental way through different changes.

Reengineering is the fundamental rethinking and radical redesign of processes to achieve dramatic improvements within critical and contemporary performance measures such as cost, quality, service and speed.[16]

First of all, it is important to remember that as previously explained in the introduction chapter (1) of this dissertation, the general goal of this implementation process and migration from the use of different tools is to improve and evolve allowing the correct development of the analytics processes. In this way, Iper will have the possibility of running in a more efficient coherence with its customers. Secondly, to understand that the information technology tools are differentiations in the way the process are facilitated.

In the necessity of improving the processes design it is important to notice that the company is taking great advantages of Qliksense, given that it is a tool that not only offers a more modern and better organized interface, most importantly, with a greater governance of the data and by the way the information is achieved it defines standards for managing those processes correctly, as shown in the Qliksense chapter (6). Iper found on CercaCodice, Twittiper and Webgate a good solution at the moment to satisfy the necessities but without ensuring the possibility of trusting in the data extracted, Right now the company is in a position of evolution and, changes in this field are now necessary, Qliksense offers not just a good solution for the management of the

analytics processes but it also represents evolution within the future because, Qliksense as a product, is also evolving and changing with the objective of offering new solutions and the implementation of this BI tool corresponds also to a change in the Iper managerial culture more data-driven and in the improvement of the internal procedures within the Iper departments where the analysis has to be carried out.

For the migration process and implementation of the tool was necessary to guarantee that all the task effectuated by the employees through the utilization of the other tree tools where developed by members of the company achieving a reduce perimeter of coverage. However, regarding to the implementation of the Qliksense solution did not required a complete re organization of the company's processes just a standardization of the processes and the inclusion of a new role. The redesign activities that took place considered the improvement of the different aspects found as problematic in the As-Is situation. The design of each of the processes permits to go according with the market changes, to maintain the company in the vanguard . Priority and time management suffered crucial changes, however always having into account the retro-compatibility of the system. The democratization of the data now is ensure and the reliability and reduction of risk is also guaranteed.

Chapter 8

Conclusions

Business intelligence tools are now a common instrument that has been internalized inside the different companies in the market, with the purpose of increasing and improving the analytics processes which are considered as a key when talking about determining future strategies, decision making processes and sales force actions. Moreover, other than increasing the current profitability of the company is also an implementation of a contingency process where the risk dismisses when is possible to anticipate, to detect and analyze in a properly way the daily facts. In particular for Iper Montebello as a large retail trade company the proper use of data will leads to the creation and opening of new markets, new strategies that allows this incumbent to mantain a competitive advantage identifying patters or clustering the clients at the end the final object, To satisfy customer expectations or create new ones is accomplished. Iper is one of the main companies interested in implementing new processes in order the create a data-driven culture where the decisions taken should be made upon the basis of trustworthy data, in especial for some areas where is fundamental to measure the improvements and the advances of a given provider or section, is also interested about gaining further benefits from the internal processes, as is described during the thesis. Consequently, BI tools are expressed as one of the most exciting possibilities for businesses in this moment and in the promising future.

During the development of the thesis were observed two big approaches, the change from a current analytics approach where the processes are made through the use of 3 different tools not guaranteed by the IT department and where the govern ability of the data is not trusty they flaws, benefits furthermore the comparison between qliksense BI tool was done, the new proposal to implement within the company creating also a new committee to evaluate the dimensions and measures by which the analysis are done and the creation of new base reports. In the elaboration was illustrated the structure, architecture, lack of standardize processes and the creation of new ones with and without the use of the BI tool. The thesis focused on the Analytics processes of Iper Montebello company, It also highlighted the strengths of using the qliksense Tool and the advantages about dismissing the use of the other 3 tools used during a

period of more than 10 years.

Along this chapter there is presented a general conclusion (8.1) of the analysis and implementation of a Business Intelligence tool in the Iper Montebello company with the idea of transforming the company in a data-driven organization, in accordance with the objectives already established, as well as the considerations that took part during the development of the process. Moreover, there are evidenced some fields for future study. (8.2).

8.1 General conclusion

The Iper Montebello Implementation generates added value to the Iper's current processes and permits to increase revenue, decrease costs, when comparing, during the thesis the practical steps and the implementation according with the current architecture based of the *DWH* conducted by the solution applied, *Qliksense* with the theoretical frameworks used in the literature review, it is possible to see that both approaches follow a similar regulation, including the resources and organizational models, results evident that there are few gaps that must be covered through some training programs and from the acquisition of different competences from external human resources.

Iper Montebello could make changes to improve its process. It is concluded from analysis that Iper should make some improvements oriented to implement the used of the BI tool within all the departments of the company so the silos can be eliminated and the data can be certificated by just one department, the idea is to create a process where the collaboration between the technicians and the savvy business resources is active and the results can be reflected on the sales and profits increment, avoiding the reprocess through a hybrid organization model. because Iper is starting to use the properties of the BI tool and the data-driven culture in order to unified the architecture and the language by which the different departments talk referring to the metrics and dimensions. In general, Iper's analytics process are not well coordinated with the theory, all the principal aspects in consideration are not aligned, starting from the goals that aren't defined by some fundamental dimensions as targets and temporal windows. It is verified that Qliksense is a valuable solution.

An incumbent in order to continue and assure the success in a market needs to follow the trends and to accomplish all the needs of the clients but, moreover it is necessary to invest in innovation and technology which allows to maintain the competitive and level among the competitors. to differentiate the company it is necessary to generate new solutions and Qliksense is one of these innovative solutions , by ensuring the analytics process to be improved, and by standardizing the source of data, increasing the efficiency, effectiveness and business decisions; taking in consideration the retail trade domain , these concepts can be applied on the development of new campaigns or the strategies that can be taken in order to penetrate in the market or in a given segment.

The analysis was concentrated on the redesign of the analytics process, the requirement process inside the company and, more specifically, from the ones carried out by the *DSI department* at first a centralized organization culture was applied. These processes define a set of procedural rules in order to standardize the activities done by several groups from different departments within the company . The main idea was to migrate these processes from the 3 solutions established priory and used during so many years without improvements, to the Qliksense Business Intelligence tool that represents a new opportunity to eliminate human errors and to reduce the manipulation of the semi structured data and increase the time focused on finding answers to the questions made to explain certain facts and models.

The advantages on the implementation of Qliksense solve these complications. The design of each of the processes is made and has become functional, with the new implementation it is modelled a system in which the actual logic is trasnformed in order to talk the same language of the Busines analyst and that can be standardize in order to continue with the development of new applications that will allow the erogation of the analytic process. From now on, the company can trust in Qliksense and the applications created and that will be created in order to satisfy all the current needs and the future requirements with data that can be investigated. The solution illustrated evidences the advantages of using new technologies and to be open to change.

8.2 **Recommendations**

According to the experience acquired during the implementation and the development of the new processes, the new BI tool, and the knowledge collected about the company . Some future considerations are proposed to take into account:

- Extend the use of the proposed Business Intelligence tool, for all the departments, this would provide better results in the analysis done by the company.
- The development and implementation of the applications within the *BI* tool must be associated and aligned if is possible with all the areas within the company this will reduce in coherency and re-work processes.
- As the idea is to reproduce a Hybrid organizational model it is necessary to include the role of business analysis inside each of the departments not just for the administration control department and the purchase one
- At some point in time, and with the future implementation of the data lake it will be possible to make deepest analysis about the data that where not considered necessary at first and will be a notorious increment of the data availability.

• Establish a methodology in order to record all the historicity of some facts as the price of the products in each sales point that will change according with the last inserted and the causes of some dis concordances can not be identified easily because of this.

Appendix A

Appendix A

A.1 Registry Table Product

[Anag_Prod]: LOAD PROD_ID, PROD_CD as ProdCod, BAR_CD as BarcodeCod, PROD_DESC as ProdDesc, PROD_CD & ' - ' & PROD_DESC as Prodotto, RET_UNIT_CD as RetUnitCod, SUBFAM_CD as SottofamigliaCod, SUBFAM_NM as SottofamigliaDesc, SUBFAM_CD & ' - ' & SUBFAM_NM as Sottofamiglia, FAM_CD as FamigliaCod, FAM_NM as FamigliaDesc, FAM_CD & ' - ' & FAM_NM as Famiglia, GRUP_CD as GruppoCod, GRUP_NM as GruppoDesc, GRUP_CD & ' - ' & GRUP_NM as Gruppo, //INSERIM_DATA_ID as DataInserimentoId, QTA_FLG as PesoPezzoFlg, //FORN_PRINC_CD as FornAnagraficoPrincCod, FORN_PRINC_ORIG_ID as FORN_ANAG_ID, IVA_VAL as ValIva, MARGINE_IMM_VAL as ValMargineImm, PR_SIZE_CD as PezzaturaCod, //PREZZO_CONS_VAL as ValPrezzoCons, //PUNTO_PREZZO_CD,

```
//UNTA_MIS_CD ,
    //PROD_TOT_CD,
    //ASTRA_CD,
    //VARIANTE_1_CD,
    //VARIANTE_2_CD,
    //COMMERCIALE_FLG,
    //REP_SET_GRUP_FAM_ID,
    //REP_SET_ID,
    REP_CD as RepartoCod,
    REP_NM as RepartoDesc,
    REP_CD & ' - ' & REP_NM as Reparto,
    SET_CD as SettoreCod,
SET_NM as SettoreDesc,
    SET_CD & ' - ' & SET_NM as Settore,
    //AGGIORN_DATA_ID,
    //RUOLO_CD,
    //BRAND_FORN_CD,
    //PROD_ORIG_CD,
    MARCHIO_TIPO_CD as MarchioTipoCod,
    //SPEC_CONSUM_CD,
    //PROD_CAUSALE_CD as CodCausale,
    MARCHIO_CD as MarchioCod,
    //PACKET_CD as CodPacket,
    //MODELLO_CD,
    //STAGIONALITA_CD as CodStagionalita,
    //COLTIVAZ_CD as CodColtivazione,
    //ORIG_NAZ_CD,
    //ORTO_CD as CodOrto,
    GARANZIA_CD as CodGaranzia,
    //STATO_CD as CodStato,
    //ORIG_REG_CD as CodRegioneOrigine,//lasciare Desc
    //STAGIONE_CD as CodStagione,//lasciare Desc
    //TAC_CD as CodTac, //lasciare Desc
    STATO_NUOVO_CD as CodStatoNuovoCod,
    //ORIG_CD,
    //TIPO_GEST_PROD_CD as CodTipoGestProdotto,
    //PROD_IMMATERIALE_CD as CodProdInmateriale,
```

//INFIAMMABILE_CD as CodInfiammabile, //CONFEZIONE_CD as CodConfezione, //PANIERE_CD as CodPaniere, //lasciare Desc PIATTAFORMA_CD as PiattaformaAnaProdCod, //TIPO_GEST_PIATT_CD, //ARTICOLO_FORNITORE_CD as CodArticoloFornitore, //DATA_NASCITA as DataNascita, //DATA_VALIDITA as DataValidita, STATO_DEPOSITO_CD as StatoDepositoAnaProdCod, //TIP0_ZUCCHERI_CD as CodTipoZuccheri, //TIPO_ALCOOL_CD as CodTipoAlcool, GRADI_ALCOOL as GradiAlcool, //DENOM_VINI_CD as CodDenomVini, //COLORE_VINI_CD as CodColoreVini, PROV_VINI_CD as ProvenienzaViniCod, //CATEG_VINI_CD as CodCategoriaVini, //GIORNI_SCADENZA, //COD_UTIF_CD, //TIPO_ASSORTIMENTO as TipoAssortimento, //Lasciare Desc //TIPO_COMP_CD as CodTipoComponente, //TIP0_PREZZ0_CD as CodTipoPrezzo, //BRAND_CD as CodBrand, //TIPO_BRAND_CD as CodTipoBrand, //FL_CROSS_DOCK as CrossDockingFlg, //FL_REVERSE_CHG, //FL_INCENTIVO, //FL_GEST_VTPRICE, //FL_BILANCIA, //FL_VENDITA_PS_PZ, //FL_VENDITA_ONLINE_CL_AND_CO, //CODICE_CATEGORIA_PREZZO as CodCategoriaPrezzo, //CODICE_ID_GRUPPO, //FL_FISSO_VARIABILE, //FL_SENZA_GLUTINE, //FL_ETNICO, FL_GEST_CELIACHIA as CeliachiaFlg,

//FL_CAPOSTIPITE,

//FL_APPROVIG_DA_SEDE,

CARTELLONE_DSC as CartelloneDesc, //FL_INDAGINE_PROD, //GRP_MERCEOLOGICO_BIL, //CODICE_PLU, //TIPO_RICORRENZA as TipoRicorrenza , UXI, UXS, UXP, ALTEZZA_PEZZO as AltezzaPezzo, LUNGHEZZA_PEZZO as LunghezzaPezzo, PROFONDITA_PEZZO as ProfonditaPezzo, PESO_MEDIO as PesoMedio, //FL_REFERENZA_PALBOX_PADRE, //FL_PRESENZA_CODICI_UNIEURO, //FL_LOCKER, PESO_SGOCCIOLATO as PesoSgocciolato, //DESCRIZIONE as DescColoreVini, //PACKET_DESC as DescPacket, //PROD_CAUSALE_DESC as DescCausaleProd, //CONFEZIONE_DESC as DescConfezione, //DESCRIZIONE as DescCategorieVini , //COLTIVAZ_DESC as DescColtivazione, GARANZIA_DESC as GaranziaDesc, MARCHIO_DESC as MarchioDesc, MARCHIO_TIPO_DESC as DescMarchioTipo, //MODELLO_DESC as DescModello, //PROD_ORIG_DESC as DescProdOrig, //ORIG_NAZ_DESC as DescNazioneOrigine, ORIG_REG_DESC as RegioneOrigineDesc, FORN_DESC AS FornDesc, //ORTO_DESC as DescOrto, //RUOLO_DESC as DescRuolo, //SPEC_CONSUM_DESC as DescSpecConsum, //STAGIONALITA_DESC as DescStagionalita, STAGIONE_DESC as StagioneDesc; //STATO_NUOVO_DESC as DescStatoNuovo, //TIPO_GEST_PROD_DESC as DescTipoGest, //PROD_IMMATERIALE_DESC as DescInmateriale,

11

```
//DESCRIZIONE as DescProvenienzaVini,
   //DESCRIZIONE as DescDeposito,
   //STATO_DESC as DescStato,
11
       TAC_DESC as TacDesc;
   //DESCRIZIONE as DescTipoAlcool,
   //PANIERE_DESC as DescPaniere,
    //UNTA_MIS_DESC as DescUnitaMisura,
    //DESCRIZIONE as DescTipoAssortimento;
SELECT
   prod.PROD_ID,
   prod.PROD_CD,
   prod.BAR_CD,
   prod.PROD_DESC,
   prod.RET_UNIT_CD,
   prod.SUBFAM_CD,
    subfam.SUBFAM_NM,
   prod.FAM_CD,
    fam.FAM_NM,
   prod.GRUP_CD,
    grup.GRUP_NM,
    //prod.INSERIM_DATA_ID,
   prod.QTA_FLG,
   //prod.FORN_PRINC_CD,
   prod.FORN_PRINC_ORIG_ID,
   prod.IVA_VAL,
   prod.MARGINE_IMM_VAL,
   prod.PR_SIZE_CD,
   //prod.PREZZO_CONS_VAL,
   //prod.PUNTO_PREZZO_CD,
   //prod.UNTA_MIS_CD,
   //prod.PROD_TOT_CD,// sono qua
   //prod.ASTRA_CD,
   //prod.VARIANTE_1_CD,
   //prod.VARIANTE_2_CD,
   //prod.COMMERCIALE_FLG,
   //prod.REP_SET_GRUP_FAM_ID,
   //prod.REP_SET_ID,
   prod.REP_CD,
```

rep.REP_NM, prod.SET_CD, sett.SET_NM, //prod.AGGIORN_DATA_ID, //prod.RUOLO_CD, //prod.BRAND_FORN_CD, //prod.PROD_ORIG_CD, prod.MARCHIO_TIPO_CD, //prod.SPEC_CONSUM_CD, //prod.PROD_CAUSALE_CD, prod.MARCHIO_CD, // prod.PACKET_CD, //prod.MODELLO_CD, //prod.STAGIONALITA_CD, //prod.COLTIVAZ_CD, //prod.ORIG_NAZ_CD, //prod.ORTO_CD, prod.GARANZIA_CD, //prod.STATO_CD, //prod.ORIG_REG_CD, //prod.STAGIONE_CD, //prod.TAC_CD, prod.STATO_NUOVO_CD, //prod.ORIG_CD, //prod.TIPO_GEST_PROD_CD, //prod.PROD_IMMATERIALE_CD, //prod.INFIAMMABILE_CD, //prod.CONFEZIONE_CD, //prod.PANIERE_CD, prod.PIATTAFORMA_CD, //prod.TIPO_GEST_PIATT_CD, //prod.ARTICOLO_FORNITORE_CD, //prod.DATA_NASCITA, //prod.DATA_VALIDITA, prod.STATO_DEPOSITO_CD, //prod.TIP0_ZUCCHERI_CD, //prod.TIP0_ALCOOL_CD, prod.GRADI_ALCOOL,

//prod.DENOM_VINI_CD, //prod.COLORE_VINI_CD, prod.PROV_VINI_CD, //prod.CATEG_VINI_CD, //prod.GIORNI_SCADENZA, //prod.COD_UTIF_CD, //prod.TIPO_ASSORTIMENTO, //prod.TIP0_COMP_CD, //prod.TIP0_PREZZ0_CD, //prod.BRAND_CD, //prod.TIPO_BRAND_CD, //prod.FL_CROSS_DOCK, //prod.FL_REVERSE_CHG, //prod.FL_INCENTIVO, //prod.FL_GEST_VTPRICE, //prod.FL_BILANCIA, //prod.FL_VENDITA_PS_PZ, //prod.FL_VENDITA_ONLINE_CL_AND_CO, //prod.CODICE_CATEGORIA_PREZZO, //prod.CODICE_ID_GRUPPO, //prod.FL_FISSO_VARIABILE, //prod.FL_SENZA_GLUTINE, //prod.FL_ETNICO, prod.FL_GEST_CELIACHIA, //prod.FL_CAPOSTIPITE, //prod.FL_APPROVIG_DA_SEDE, prod.CARTELLONE_DSC, //prod.FL_INDAGINE_PROD, //prod.GRP_MERCEOLOGICO_BIL, //prod.CODICE_PLU, //prod.TIPO_RICORRENZA, prod.UXI, prod.UXS, prod.UXP, prod.ALTEZZA_PEZZO, prod.LUNGHEZZA_PEZZO, prod.PROFONDITA_PEZZO, prod.PESO_MEDIO,

//prod.FL_REFERENZA_PALBOX_PADRE, //prod.FL_PRESENZA_CODICI_UNIEURO, //prod.FL_LOCKER, prod.PESO_SGOCCIOLATO, //colorevini.DESCRIZIONE, //catevini.DESCRIZIONE, //packet.PACKET_DESC, //causaleprod.PROD_CAUSALE_DESC, //confezione.CONFEZIONE_DESC, //paniere.PANIERE_DESC, //coltivaz.COLTIVAZ_DESC, garanzia.GARANZIA_DESC, marchio.MARCHIO_DESC, marchiotipo.MARCHIO_TIPO_DESC, //modello.MODELLO_DESC, //orig.PROD_ORIG_DESC, //naz.ORIG_NAZ_DESC, reg.ORIG_REG_DESC, //orto.ORTO_DESC, //ruolo.RUOLO_DESC, //spec.SPEC_CONSUM_DESC, //stagionalita.STAGIONALITA_DESC, stagione.STAGIONE_DESC, //statonuovo.STATO_NUOVO_DESC, //tipogest.TIPO_GEST_PROD_DESC, //tipoinmater.PROD_IMMATERIALE_DESC, //provini.DESCRIZIONE, //deposito.DESCRIZIONE, //stato.STATO_DESC, forn.FORN_DESC, tac.TAC_DESC //alcool.DESCRIZIONE,

//zuccheri.DESCRIZIONE,

11

//unitamis.UNTA_MIS_DESC,

//tipoasso.DESCRIZIONE

FROM FIN_DATA.DIM_PROD prod

LEFT JOIN FIN_DATA.ANA_REP rep on prod.REP_CD=rep.REP_CD LEFT join FIN_DATA.ANA_SET sett on prod.SET_CD=sett.SET_CD and prod.REP_CD=sett.REP_CD LEFT JOIN FIN_DATA.ANA_GRUP grup on prod.GRUP_CD=grup.GRUP_CD and prod.SET_CD=grup.SET_CD and prod.REP_CD=grup.REP_CD LEFT Join FIN_DATA.ANA_FAM fam on prod.FAM_CD=fam.FAM_CD and prod.GRUP_CD=fam.GRUP_CD and prod.SET_CD=fam.SET_CD and prod.REP_CD=fam.REP_CD LEFT JOIN FIN_DATA.ANA_SUBFAM subfam on prod.SUBFAM_CD= subfam.SUBFAM_CD and prod.FAM_CD=subfam.FAM_CD and prod.GRUP_CD=subfam.GRUP_CD and prod.SET_CD=subfam.SET_CD and prod.REP_CD=subfam.REP_CD // LEFT JOIN FIN_DATA.DIM_FORN forn on prod.FORN_PRINC_ORIG_ID=forn.FORN_ID //LEFT JOIN FIN_DATA.ANA_CATEGORIE_VINI catevini on prod.CATEG_VINI_CD= catevini.CATEG_VINI_CD //LEFT JOIN FIN_DATA.ANA_COLORE_VINI colorevini on prod.COLORE_VINI_CD= colorevini.COLORE_VINI_CD //LEFT JOIN FIN_DATA.ANA_PACKET packet on prod.PACKET_CD=packet.PACKET_CD //LEFT JOIN FIN_DATA.ANA_PROD_CAUSALE causaleprod on prod.PROD_CAUSALE_CD= causaleprod.PROD_CAUSALE_CD //LEFT JOIN FIN_DATA.ANA_PROD_COD_CONFEZIONE confezione on prod.CONFEZIONE_CD= confezione.CONFEZIONE_CD //LEFT JOIN FIN_DATA.ANA_PROD_COD_PANIERE paniere on prod.PANIERE_CD = paniere.PANIERE_CD //LEFT JOIN FIN_DATA.ANA_PROD_COLTIVAZ coltivaz on prod.COLTIVAZ_CD= coltivaz.COLTIVAZ_CD LEFT JOIN FIN_DATA.ANA_PROD_GARANZIA garanzia on prod.GARANZIA_CD=garanzia.GARANZIA_CD LEFT JOIN FIN_DATA.ANA_PROD_MARCHIO marchio on prod.MARCHIO_CD=marchio.MARCHIO_CD LEFT JOIN FIN_DATA.ANA_PROD_MARCHIO_TIPO marchiotipo on prod.MARCHIO_TIPO_CD=marchiotipo.MARCHIO_TIPO_CD //LEFT JOIN FIN_DATA.ANA_PROD_MODELLO modello on prod.MODELLO_CD=modello.MODELLO_CD //LEFT JOIN FIN_DATA.ANA_PROD_ORIG orig on prod.PROD_ORIG_CD= orig.PROD_ORIG_CD

//LEFT JOIN FIN_DATA.ANA_PROD_ORIG_NAZ naz on prod.ORIG_NAZ_CD= naz.ORIG_NAZ_CD LEFT JOIN FIN_DATA.ANA_PROD_ORIG_REG reg on prod.ORIG_REG_CD= reg.ORIG_REG_CD //LEFT JOIN FIN_DATA.ANA_PROD_ORTO orto on prod.ORTO_CD= orto.ORTO_CD //LEFT JOIN FIN_DATA.ANA_PROD_RUOLO ruolo on prod.RUOLO_CD= ruolo.RUOLO_CD //LEFT JOIN FIN_DATA.ANA_PROD_SPEC_CONSUM spec on prod.SPEC_CONSUM_CD= spec.SPEC_CONSUM_CD //LEFT JOIN FIN_DATA.ANA_PROD_STAGIONALITA stagionalita on prod.STAGIONALITA_CD=stagionalita.STAGIONALITA_CD LEFT JOIN FIN_DATA.ANA_PROD_STAGIONE stagione on prod.STAGIONE_CD= stagione.STAGIONE_CD //LEFT JOIN FIN_DATA.ANA_PROD_STATO_NUOVO statonuovo on prod.STATO_NUOVO_CD=statonuovo.STATO_NUOVO_CD //LEFT JOIN FIN_DATA.ANA_PROD_TIPO_GEST tipogest on prod.TIPO_GEST_PROD_CD= tipogest.TIPO_GEST_PROD_CD //LEFT JOIN FIN_DATA.ANA_PROD_TIPO_IMMATER tipoinmater on prod.PROD_IMMATERIALE_CD= tipoinmater.PROD_IMMATERIALE_CD //LEFT JOIN FIN_DATA.ANA_PROVENIENZA_VINI provini on prod.PROV_VINI_CD=provini.PROV_VINI_CD //LEFT JOIN FIN_DATA.ANA_STATI_DEPOSITO deposito on prod.STATO_DEPOSITO_CD =deposito.STATO_DEPOSITO_CD //LEFT JOIN FIN_DATA.ANA_STATO stato on prod.STATO_CD=stato.STATO_CD LEFT JOIN FIN_DATA.ANA_TAC tac on prod.TAC_CD=tac.TAC_CD; //LEFT JOIN FIN_DATA.ANA_TIPO_ALCOOL alcool on prod.TIPO_ALCOOL_CD= alcool.TIPO_ALCOOL_CD //LEFT JOIN FIN_DATA.ANA_TIPO_ZUCCHERI zuccheri on prod.TIPO_ZUCCHERI_CD= zuccheri.TIPO_ZUCCHERI_CD //LEFT JOIN FIN_DATA.DIM_UNITA_MISURA unitamis on prod.UNTA_MIS_CD=unitamis.UNTA_MIS_CD //LEFT JOIN FIN_DATA.ANA_TIPO_ASSSORTIMENTO tipoasso on prod.TIPO_ASSORTIMENTO_CD= tipoasso.TIPO_ASSORTIMENTO_CD; STORE Anag_Prod INTO [Lib://QVD_folder (sense_digital_innovation1)/Qvd_AnagProd.QVD](qvd);

A.2 Registry Table Providers

```
[Anag_Prod]:
LOAD PROD_ID,
PROD_CD as ProdCod,
BAR_CD as BarcodeCod,
PROD_DESC as ProdDesc,
    PROD_CD & ' - ' & PROD_DESC as Prodotto,
RET_UNIT_CD as RetUnitCod,
SUBFAM_CD as SottofamigliaCod,
    SUBFAM_NM as SottofamigliaDesc,
    SUBFAM_CD & ' - ' & SUBFAM_NM as Sottofamiglia,
FAM_CD as FamigliaCod,
   FAM_NM as FamigliaDesc,
   FAM_CD & ' - ' & FAM_NM as Famiglia,
GRUP_CD as GruppoCod,
   GRUP_NM as GruppoDesc,
   GRUP_CD & ' - ' & GRUP_NM as Gruppo,
//INSERIM_DATA_ID as DataInserimentoId,
QTA_FLG as PesoPezzoFlg,
//FORN_PRINC_CD as FornAnagraficoPrincCod,
FORN_PRINC_ORIG_ID as FORN_ANAG_ID,
IVA_VAL as Vallva,
MARGINE_IMM_VAL as ValMargineImm,
   PR_SIZE_CD as PezzaturaCod,
   //PREZZO_CONS_VAL as ValPrezzoCons,
   //PUNTO_PREZZO_CD,
   //UNTA_MIS_CD ,
   //PROD_TOT_CD,
   //ASTRA_CD,
   //VARIANTE_1_CD,
    //VARIANTE_2_CD,
   //COMMERCIALE_FLG,
    //REP_SET_GRUP_FAM_ID,
    //REP_SET_ID,
   REP_CD as RepartoCod,
   REP_NM as RepartoDesc,
   REP_CD & ' - ' & REP_NM as Reparto,
    SET_CD as SettoreCod,
```

SET_NM as SettoreDesc,

SET_CD & ' - ' & SET_NM as Settore,

//AGGIORN_DATA_ID,

//RUOLO_CD,

//BRAND_FORN_CD,

//PROD_ORIG_CD,

MARCHIO_TIPO_CD as MarchioTipoCod,

//SPEC_CONSUM_CD,

//PROD_CAUSALE_CD as CodCausale,

MARCHIO_CD as MarchioCod,

//PACKET_CD as CodPacket,

//MODELLO_CD ,

//STAGIONALITA_CD as CodStagionalita,

//COLTIVAZ_CD as CodColtivazione,

//ORIG_NAZ_CD,

//ORTO_CD as CodOrto,

GARANZIA_CD as CodGaranzia,

//STATO_CD as CodStato,

//ORIG_REG_CD as CodRegioneOrigine,//lasciare Desc

//STAGIONE_CD as CodStagione,//lasciare Desc

//TAC_CD as CodTac, //lasciare Desc

STATO_NUOVO_CD as CodStatoNuovoCod,

//ORIG_CD,

//TIPO_GEST_PROD_CD as CodTipoGestProdotto,

//PROD_IMMATERIALE_CD as CodProdInmateriale,

//INFIAMMABILE_CD as CodInfiammabile,

//CONFEZIONE_CD as CodConfezione,

//PANIERE_CD as CodPaniere, //lasciare Desc

PIATTAFORMA_CD as PiattaformaAnaProdCod,

//TIPO_GEST_PIATT_CD,

//ARTICOLO_FORNITORE_CD as CodArticoloFornitore,

//DATA_NASCITA as DataNascita,

//DATA_VALIDITA as DataValidita,

STATO_DEPOSITO_CD as StatoDepositoAnaProdCod,

//TIP0_ZUCCHERI_CD as CodTipoZuccheri,

//TIPO_ALCOOL_CD as CodTipoAlcool,

GRADI_ALCOOL as GradiAlcool,

//DENOM_VINI_CD as CodDenomVini,

//COLORE_VINI_CD as CodColoreVini, PROV_VINI_CD as ProvenienzaViniCod, //CATEG_VINI_CD as CodCategoriaVini, //GIORNI_SCADENZA, //COD_UTIF_CD, //TIPO_ASSORTIMENTO as TipoAssortimento, //Lasciare Desc //TIPO_COMP_CD as CodTipoComponente, //TIPO_PREZZO_CD as CodTipoPrezzo, //BRAND_CD as CodBrand, //TIPO_BRAND_CD as CodTipoBrand, //FL_CROSS_DOCK as CrossDockingFlg, //FL_REVERSE_CHG, //FL_INCENTIVO, //FL_GEST_VTPRICE, //FL_BILANCIA, //FL_VENDITA_PS_PZ, //FL_VENDITA_ONLINE_CL_AND_CO, //CODICE_CATEGORIA_PREZZO as CodCategoriaPrezzo, //CODICE_ID_GRUPPO, //FL_FISSO_VARIABILE, //FL_SENZA_GLUTINE, //FL_ETNICO, FL_GEST_CELIACHIA as CeliachiaFlg, //FL_CAPOSTIPITE, //FL_APPROVIG_DA_SEDE, CARTELLONE_DSC as CartelloneDesc, //FL_INDAGINE_PROD, //GRP_MERCEOLOGICO_BIL, //CODICE_PLU, //TIPO_RICORRENZA as TipoRicorrenza , UXI, UXS, UXP, ALTEZZA_PEZZO as AltezzaPezzo, LUNGHEZZA_PEZZO as LunghezzaPezzo, PROFONDITA_PEZZO as ProfonditaPezzo, PESO_MEDIO as PesoMedio, //FL_REFERENZA_PALBOX_PADRE,

//FL_PRESENZA_CODICI_UNIEURO, //FL_LOCKER, PESO_SGOCCIOLATO as PesoSgocciolato, //DESCRIZIONE as DescColoreVini, //PACKET_DESC as DescPacket, //PROD_CAUSALE_DESC as DescCausaleProd, //CONFEZIONE_DESC as DescConfezione, //DESCRIZIONE as DescCategorieVini , //COLTIVAZ_DESC as DescColtivazione, GARANZIA_DESC as GaranziaDesc, MARCHIO_DESC as MarchioDesc, MARCHIO_TIPO_DESC as DescMarchioTipo, //MODELLO_DESC as DescModello, //PROD_ORIG_DESC as DescProdOrig, //ORIG_NAZ_DESC as DescNazioneOrigine, ORIG_REG_DESC as RegioneOrigineDesc, FORN_DESC AS FornDesc, //ORTO_DESC as DescOrto, //RUOLO_DESC as DescRuolo, //SPEC_CONSUM_DESC as DescSpecConsum,

//STAGIONALITA_DESC as DescStagionalita,

STAGIONE_DESC as StagioneDesc;

//STATO_NUOVO_DESC as DescStatoNuovo,

//TIPO_GEST_PROD_DESC as DescTipoGest,

//PROD_IMMATERIALE_DESC as DescInmateriale,

//DESCRIZIONE as DescProvenienzaVini,

//DESCRIZIONE as DescDeposito,

//STATO_DESC as DescStato,

11 TAC_DESC as TacDesc;

//DESCRIZIONE as DescTipoAlcool,

//PANIERE_DESC as DescPaniere,

//UNTA_MIS_DESC as DescUnitaMisura,

//DESCRIZIONE as DescTipoAssortimento;

SELECT

11

prod.PROD_ID, prod.PROD_CD, prod.BAR_CD, prod.PROD_DESC,

prod.RET_UNIT_CD, prod.SUBFAM_CD, subfam.SUBFAM_NM, prod.FAM_CD, fam.FAM_NM, prod.GRUP_CD, grup.GRUP_NM, //prod.INSERIM_DATA_ID, prod.QTA_FLG, //prod.FORN_PRINC_CD, prod.FORN_PRINC_ORIG_ID, prod.IVA_VAL, prod.MARGINE_IMM_VAL, prod.PR_SIZE_CD, //prod.PREZZO_CONS_VAL, //prod.PUNTO_PREZZO_CD, //prod.UNTA_MIS_CD, //prod.PROD_TOT_CD,// sono qua //prod.ASTRA_CD, //prod.VARIANTE_1_CD, //prod.VARIANTE_2_CD, //prod.COMMERCIALE_FLG, //prod.REP_SET_GRUP_FAM_ID, //prod.REP_SET_ID, prod.REP_CD, rep.REP_NM, prod.SET_CD, sett.SET_NM, //prod.AGGIORN_DATA_ID, //prod.RUOLO_CD, //prod.BRAND_FORN_CD, //prod.PROD_ORIG_CD, prod.MARCHIO_TIPO_CD, //prod.SPEC_CONSUM_CD, //prod.PROD_CAUSALE_CD, prod.MARCHIO_CD, // prod.PACKET_CD, //prod.MODELLO_CD,

//prod.STAGIONALITA_CD, //prod.COLTIVAZ_CD, //prod.ORIG_NAZ_CD, //prod.ORTO_CD, prod.GARANZIA_CD, //prod.STAT0_CD, //prod.ORIG_REG_CD, //prod.STAGIONE_CD, //prod.TAC_CD, prod.STATO_NUOVO_CD, //prod.ORIG_CD, //prod.TIPO_GEST_PROD_CD, //prod.PROD_IMMATERIALE_CD, //prod.INFIAMMABILE_CD, //prod.CONFEZIONE_CD, //prod.PANIERE_CD, prod.PIATTAFORMA_CD, //prod.TIPO_GEST_PIATT_CD, //prod.ARTICOLO_FORNITORE_CD, //prod.DATA_NASCITA, //prod.DATA_VALIDITA, prod.STATO_DEPOSITO_CD, //prod.TIP0_ZUCCHERI_CD, //prod.TIP0_ALCOOL_CD, prod.GRADI_ALCOOL, //prod.DENOM_VINI_CD, //prod.COLORE_VINI_CD, prod.PROV_VINI_CD, //prod.CATEG_VINI_CD, //prod.GIORNI_SCADENZA, //prod.COD_UTIF_CD, //prod.TIPO_ASSORTIMENTO, //prod.TIPO_COMP_CD, //prod.TIP0_PREZZ0_CD, //prod.BRAND_CD, //prod.TIPO_BRAND_CD, //prod.FL_CROSS_DOCK, //prod.FL_REVERSE_CHG,

//prod.FL_INCENTIVO, //prod.FL_GEST_VTPRICE, //prod.FL_BILANCIA, //prod.FL_VENDITA_PS_PZ, //prod.FL_VENDITA_ONLINE_CL_AND_CO, //prod.CODICE_CATEGORIA_PREZZO, //prod.CODICE_ID_GRUPPO, //prod.FL_FISSO_VARIABILE, //prod.FL_SENZA_GLUTINE, //prod.FL_ETNICO, prod.FL_GEST_CELIACHIA, //prod.FL_CAPOSTIPITE, //prod.FL_APPROVIG_DA_SEDE, prod.CARTELLONE_DSC, //prod.FL_INDAGINE_PROD, //prod.GRP_MERCEOLOGICO_BIL, //prod.CODICE_PLU, //prod.TIPO_RICORRENZA, prod.UXI, prod.UXS, prod.UXP, prod.ALTEZZA_PEZZO, prod.LUNGHEZZA_PEZZO, prod.PROFONDITA_PEZZO, prod.PESO_MEDIO, //prod.FL_REFERENZA_PALBOX_PADRE, //prod.FL_PRESENZA_CODICI_UNIEURO, //prod.FL_LOCKER, prod.PESO_SGOCCIOLATO, //colorevini.DESCRIZIONE, //catevini.DESCRIZIONE, //packet.PACKET_DESC, //causaleprod.PROD_CAUSALE_DESC, //confezione.CONFEZIONE_DESC, //paniere.PANIERE_DESC, //coltivaz.COLTIVAZ_DESC, garanzia.GARANZIA_DESC,

marchio.MARCHIO_DESC,

marchiotipo.MARCHIO_TIPO_DESC, //modello.MODELLO_DESC, //orig.PROD_ORIG_DESC, //naz.ORIG_NAZ_DESC, reg.ORIG_REG_DESC, //orto.ORTO_DESC, //ruolo.RUOLO_DESC, //spec.SPEC_CONSUM_DESC, //stagionalita.STAGIONALITA_DESC, stagione.STAGIONE_DESC, //statonuovo.STATO_NUOVO_DESC, //tipogest.TIPO_GEST_PROD_DESC, //tipoinmater.PROD_IMMATERIALE_DESC, //provini.DESCRIZIONE, //deposito.DESCRIZIONE, //stato.STATO_DESC, forn.FORN_DESC, tac.TAC_DESC

//alcool.DESCRIZIONE,

11

//zuccheri.DESCRIZIONE,

//unitamis.UNTA_MIS_DESC,

//tipoasso.DESCRIZIONE

FROM FIN_DATA.DIM_PROD prod

LEFT JOIN FIN_DATA.ANA_REP rep on prod.REP_CD=rep.REP_CD

```
LEFT join FIN_DATA.ANA_SET sett on prod.SET_CD=sett.SET_CD
and prod.REP_CD=sett.REP_CD
LEFT JOIN FIN_DATA.ANA_GRUP grup on prod.GRUP_CD=grup.GRUP_CD
and prod.SET_CD=grup.SET_CD and prod.REP_CD=grup.REP_CD
LEFT Join FIN_DATA.ANA_FAM fam on prod.FAM_CD=fam.FAM_CD and
prod.GRUP_CD=fam.GRUP_CD and prod.SET_CD=fam.SET_CD and
prod.REP_CD=fam.REP_CD
LEFT JOIN FIN_DATA.ANA_SUBFAM subfam on prod.SUBFAM_CD=
subfam.SUBFAM_CD and prod.FAM_CD=subfam.FAM_CD and
prod.GRUP_CD=subfam.GRUP_CD and prod.SET_CD=subfam.SET_CD and
```

A.2. REGISTRY TABLE PROVIDERS

// LEFT JOIN FIN_DATA.DIM_FORN forn on prod.FORN_PRINC_ORIG_ID=forn.FORN_ID //LEFT JOIN FIN_DATA.ANA_CATEGORIE_VINI catevini on prod.CATEG_VINI_CD= catevini.CATEG_VINI_CD //LEFT JOIN FIN_DATA.ANA_COLORE_VINI colorevini on prod.COLORE_VINI_CD= colorevini.COLORE_VINI_CD //LEFT JOIN FIN_DATA.ANA_PACKET packet on prod.PACKET_CD=packet.PACKET_CD //LEFT JOIN FIN_DATA.ANA_PROD_CAUSALE causaleprod on prod.PROD_CAUSALE_CD= causaleprod.PROD_CAUSALE_CD //LEFT JOIN FIN_DATA.ANA_PROD_COD_CONFEZIONE confezione on prod.CONFEZIONE_CD= confezione.CONFEZIONE_CD //LEFT JOIN FIN_DATA.ANA_PROD_COD_PANIERE paniere on prod.PANIERE_CD = paniere.PANIERE_CD //LEFT JOIN FIN_DATA.ANA_PROD_COLTIVAZ coltivaz on prod.COLTIVAZ_CD= coltivaz.COLTIVAZ_CD LEFT JOIN FIN_DATA.ANA_PROD_GARANZIA garanzia on prod.GARANZIA_CD=garanzia.GARANZIA_CD LEFT JOIN FIN_DATA.ANA_PROD_MARCHIO marchio on prod.MARCHIO_CD=marchio.MARCHIO_CD LEFT JOIN FIN_DATA.ANA_PROD_MARCHIO_TIPO marchiotipo on prod.MARCHIO_TIPO_CD=marchiotipo.MARCHIO_TIPO_CD //LEFT JOIN FIN_DATA.ANA_PROD_MODELLO modello on prod.MODELLO_CD=modello.MODELLO_CD //LEFT JOIN FIN_DATA.ANA_PROD_ORIG orig on prod.PROD_ORIG_CD= orig.PROD_ORIG_CD //LEFT JOIN FIN_DATA.ANA_PROD_ORIG_NAZ naz on prod.ORIG_NAZ_CD= naz.ORIG_NAZ_CD LEFT JOIN FIN_DATA.ANA_PROD_ORIG_REG reg on prod.ORIG_REG_CD= reg.ORIG_REG_CD //LEFT JOIN FIN_DATA.ANA_PROD_ORTO orto on prod.ORTO_CD= orto.ORTO_CD //LEFT JOIN FIN_DATA.ANA_PROD_RUOLO ruolo on prod.RUOLO_CD= ruolo.RUOLO_CD //LEFT JOIN FIN_DATA.ANA_PROD_SPEC_CONSUM spec on prod.SPEC_CONSUM_CD= spec.SPEC_CONSUM_CD //LEFT JOIN FIN_DATA.ANA_PROD_STAGIONALITA stagionalita on prod.STAGIONALITA_CD=stagionalita.STAGIONALITA_CD LEFT JOIN FIN_DATA.ANA_PROD_STAGIONE stagione on

prod.STAGIONE_CD= stagione.STAGIONE_CD //LEFT JOIN FIN_DATA.ANA_PROD_STATO_NUOVO statonuovo on prod.STATO_NUOVO_CD=statonuovo.STATO_NUOVO_CD //LEFT JOIN FIN_DATA.ANA_PROD_TIPO_GEST tipogest on prod.TIPO_GEST_PROD_CD= tipogest.TIPO_GEST_PROD_CD //LEFT JOIN FIN_DATA.ANA_PROD_TIPO_IMMATER tipoinmater on prod.PROD_IMMATERIALE_CD= tipoinmater.PROD_IMMATERIALE_CD //LEFT JOIN FIN_DATA.ANA_PROVENIENZA_VINI provini on prod.PROV_VINI_CD=provini.PROV_VINI_CD //LEFT JOIN FIN_DATA.ANA_STATI_DEPOSITO deposito on prod.STATO_DEPOSITO_CD =deposito.STATO_DEPOSITO_CD //LEFT JOIN FIN_DATA.ANA_STATO stato on prod.STATO_CD=stato.STATO_CD LEFT JOIN FIN_DATA.ANA_TAC tac on prod.TAC_CD=tac.TAC_CD; //LEFT JOIN FIN_DATA.ANA_TIPO_ALCOOL alcool on prod.TIPO_ALCOOL_CD= alcool.TIPO_ALCOOL_CD //LEFT JOIN FIN_DATA.ANA_TIPO_ZUCCHERI zuccheri on prod.TIPO_ZUCCHERI_CD= zuccheri.TIPO_ZUCCHERI_CD //LEFT JOIN FIN_DATA.DIM_UNITA_MISURA unitamis on prod.UNTA_MIS_CD=unitamis.UNTA_MIS_CD //LEFT JOIN FIN_DATA.ANA_TIPO_ASSSORTIMENTO tipoasso on prod.TIPO_ASSORTIMENTO_CD= tipoasso.TIPO_ASSORTIMENTO_CD; STORE Anag_Prod INTO [Lib://QVD_folder (sense_digital_innovation1)/Qvd_AnagProd.QVD](qvd);

A.3 Registry Table Sales Stores

```
[Anag_Enti]:
```

LOAD

```
//GEO_ID,
ENTE_ID,
ENTE_CD as EnteCod,
ENTE_TIPO_CD as TipoEnteCod,
ENTE_DESC as EnteDesc,
//REPLICA_DATA_ID,
//ENTE_ACR_CD,
//AREA_CD,
//INSERIM_DATA_ID,
```

//AGGIORN_DATA_ID,
//FLAG_LOAD_ACQ,
AREA_DESC as AreaDesc;

Select

//	ente.GEO_ID,	
	ente.ENTE_ID,	
	ente.ENTE_CD,	
	ente.ENTE_TIPO_CD,	
	ente.ENTE_DESC,	
//	ente.REPLICA_DATA_ID,	
//	ente.ENTE_ACR_CD,	
//	ente.AREA_CD,	
//	//ente.INSERIM_DATA_ID,	
//	//ente.AGGIORN_DATA_ID,	
//	ente.FLAG_LOAD_ACQ,	
	area.AREA_DESC	
FROM	FIN_DATA.DIM_ENTI ente	
LEFT	JOIN FIN_DATA.ANA_AREE area on ente.AREA_CD=area.AREA_CD;	
STORE	E Anag_Enti INTO [Lib://QVD_folder	
(sens	<pre>se_digital_innovation1)/Qvd_AnagEnte.QVD](qvd);</pre>	

A.4 Registry Table Calendar

```
Periodi:
LOAD
     //DATA_CD as Datacod,
     ANNO_ID as Anno,
     Right(TRIMESTRE_ID,1) as Trimestre,
     Right(MESE_ID,2) as MeseNum,
     Month(Date(Date#(PERIOD_ID,'YYYYMMDD'),'DD/MM/YYYY')) as MeseNome,
     Right(SETTIMANA_ID,2) as Settimana,
     //DATA_DESC as DataDesc,
     PERIOD_ID,
     Date(Date#(PERIOD_ID,'YYYYMMDD'),'DD/MM/YYYY') as Giorno,
     //LY_DATA_ID,
     Right(SEMESTRE_ID,1) as Semestre,
```

```
Right(MICROSETTIMANA_ID,3) as Microsettimana,
      Capitalize(GIORNO_CD) as GiornoSettimana,
      Month(Date(Date#(PERIOD_ID,'YYYYMMDD'),'DD/MM/YYYY'))
      & '/' & ANNO_ID as MeseAnno,
      Right(SETTIMANA_ID,2)& '/' & ANNO_ID as SettimanaAnno
      ;
      //PERIOD_ID_AP;
SELECT
      //DATA_CD,
      ANNO_ID,
      TRIMESTRE_ID,
      MESE_ID,
      SETTIMANA_ID,
      //DATA_DESC,
      PERIOD_ID,
      //LY_DATA_ID,
      SEMESTRE_ID,
      MICROSETTIMANA_ID,
      GIORNO_CD
      //PERIOD_ID_AP
FROM FIN_DATA.DIM_PERIODI
WHERE PERIOD_ID>=$(vStartDate) and PERIOD_ID
<=(select GIORNO_VEND from FIN_DATA.V_SYS_PARAMETRI);
STORE Periodi INTO [Lib://QVD_folder
(sense_digital_innovation1)/Qvd_Periodi.qvd](qvd);
```

A.5 Fact Table Assortment PDV

```
[AssortimentoPdv]:
LOAD
//ENTE_CD as EnteCod,
ENTE_ID,
//PROD_CD as ProdCod,
PROD_ID,
```

```
// TIPO_ETICH,
```

```
// TIPO_GEST_STK,
    QTA_ULT_IMMESSO as QtaUltimoImm,
```

ULT_IMMESSO as UltimoImm,

DT_ULTIMO_IMMESSO as DataUltimoImm,

- // IMPORTANZA,
- // PANIERE_LOCALE,
- // PANIERE_COLORE,
- // MARGINE_FISSO,
- // DT_INS_ASS,
- // DT_FUORI_ASS,
 - DT_ULT_VENDITA as DataUltimaVendita,
- // FACING,
- // PROFONDITA,
- // COSTO_ORIG_SDC,
- // PRZ_ORIG_SDC,
- // COD_COLLEZIONE,
- // DATA_ELAB,
 - //FORN_PRINC as FORN_ANAG_CD, FORN_ID_PRINC as FORN_ANAG_ID, FORN_ID_PRINC_ACQ,
- // FORN_PRINC_ACQ as FornPrincAcqCod,
- // ENTE_DESC as EnteDesc,
- // FORN_DESC AS FornDesc,
- // PROD_DESC as ProdDesc,
 - PERIOD_ID as DataPrezzi,

PROMO_ID,

VAL_PRZ_ORDINARIO as PrezzoOrdinario,

- VAL_PRZ_VEN as PrezzoVendita,
- VAL_PRZ_NET as PrezzoVenditaNetto,
- VAL_PRZ_VEN_SCONTATO as PrezzoScontato

;

SELECT

- // prodpdv.ENTE_CD,
 prodpdv.ENTE_ID,
- // prodpdv.PROD_CD,
 prodpdv.PROD_ID,
- // prodpdv.TIPO_ETICH,

11 prodpdv.TIPO_GEST_STK, prodpdv.QTA_ULT_IMMESSO, prodpdv.ULT_IMMESSO, prodpdv.DT_ULTIMO_IMMESSO, 11 prodpdv.IMPORTANZA, // prodpdv.PANIERE_LOCALE, 11 prodpdv.PANIERE_COLORE, 11 prodpdv.MARGINE_FISSO, // prodpdv.DT_INS_ASS, 11 prodpdv.DT_FUORI_ASS, prodpdv.DT_ULT_VENDITA, // prodpdv.FACING, 11 prodpdv.PROFONDITA, 11 prodpdv.COSTO_ORIG_SDC, 11 prodpdv.PRZ_ORIG_SDC, 11 prodpdv.COD_COLLEZIONE, 11 prodpdv.DATA_ELAB, // prodpdv.FORN_PRINC prezzipdv.PERIOD_ID, prodpdv.FORN_ID_PRINC, prodpdv.UXI, prodpdv.UXS, prodpdv.UXP, prodpdv.FORN_ID_PRINC_ACQ, prezzipdv.PROMO_ID, prezzipdv.VAL_PRZ_ORDINARIO, prezzipdv.VAL_PRZ_VEN, prezzipdv.VAL_PRZ_NET, prezzipdv.VAL_PRZ_VEN_SCONTATO //prodpdv.FORN_PRINC_ACQ FROM FIN_DATA.DIM_PROD_PDV prodpdv join FIN_DATA.DIM_PREZZI_PDV prezzipdv on prodpdv.ENTE_ID=prezzipdv.ENTE_ID and prodpdv.PROD_ID=prezzipdv.PROD_ID and prezzipdv.PERIOD_ID = (select GIORNO_VEND from FIN_DATA.V_SYS_PARAMETRI); STORE AssortimentoPdv INTO [Lib://QVD_folder (sense_digital_innovation1)/Qvd_AssortimentoPdv.qvd](qvd);

A.6 Fact Table Daily Sales

VendutoPromoGiorno:

Load

PROMO_VEN_QTA_PS as PromoVenQtaPs, PROMO_VEN_QTA_PZ as PromoVenQtaPz, PROMO_VEN_VAL as PromoVenVal, PROMO_VEN_UTILE AS PromoVenUtile, PROMO_VEN_UTILE/PROMO_VEN_VAL as MarginePromo1%, PROMO_VEN_NET_VAL as PromoVenNetVal, PROMO_COSTO_VAL as PromoCostoVal, //REPLICA_ID , PROMO_ID, PERIOD_ID, ENTE_ID, PROD_ID, FL_NSDR as NsdrFlg, FL_OK as OkFlg, LIVELLO_OR_CD as LivelloOrCod, //PROMO_COSTO_VAL_ORIG as PromoCostValOrig , //FL_IPER_SPORT as IperSportFlg, PROMO_ULTIMO_COSTO as PromoUltimoCosto, PERIOD_ID & ENTE_ID & PROD_ID as PERIOD_ENTE_PROD_KEY;

Select

venprom.ENTE_CD, venprom.PROD_CD, venprom.RET_UNIT_CD, venprom.PROMO_VEN_QTA_PS, venprom.PROMO_VEN_QTA_PS, venprom.PROMO_VEN_QTA_PZ, venprom.PROMO_VEN_UTILE, venprom.PROMO_VEN_UTILE, venprom.PROMO_VEN_NET_VAL, venprom.PROMO_COSTO_VAL,

```
//venprom.REPLICA_ID,
venprom.PROMO_ID,
venprom.PERIOD_ID,
venprom.ENTE_ID,
venprom.PROD_ID,
venprom.FL_NSDR,
venprom.FL_OK,
venprom.LIVELLO_OR_CD,
//venprom.PROMO_COSTO_VAL_ORIG,
venprom.PROMO_ULTIMO_COSTO
//venprom.FL_IPER_SPORT
```

FROM FIN_DATA.FACT_VEN_PROMO_GIORNO venprom

```
where PERIOD_ID>=$(vStartDate);
STORE VendutoPromoGiorno INTO [Lib://QVD_folder
(sense_digital_innovation1)/Qvd_VendutoPromoGiorno.QVD](qvd);
```

A.7 Registry Table Merchandise Structure

```
[StrutturaMerceologica]:
LOAD
    REP_CD as RepartoCod,
REP_NM as RepartoDesc,
    REP_CD & ' - ' & REP_NM as Reparto,
SET_CD as SettoreCod,
SET_NM as SettoreDesc,
    SET_CD & ' - ' & SET_NM as Settore,
GRUP_CD as GruppoCod,
GRUP_NM as GruppoDesc,
    GRUP_CD & ' - ' & GRUP_NM as Gruppo,
FAM_CD as FamigliaCod,
FAM_NM as FamigliaDesc,
    FAM_CD & ' - ' & FAM_NM as Famiglia,
SUBFAM_CD as SottofamigliaCod,
SUBFAM_NM as SottofamigliaDesc,
```

```
SUBFAM_CD & ' - ' & SUBFAM_NM as Sottofamiglia,
   SET_CD & GRUP_CD & FAM_CD & SUBFAM_CD as RetUnit
;
SELECT
    subfam.REP_CD,
rep.REP_NM,
subfam.SET_CD,
sett.SET_NM,
subfam.GRUP_CD,
grup.GRUP_NM,
subfam.FAM_CD,
fam.FAM_NM,
subfam.SUBFAM_CD,
subfam.SUBFAM_NM
FROM FIN_DATA.ANA_SUBFAM subfam
join FIN_DATA.ANA_FAM fam on subfam.FAM_CD=fam.FAM_CD
and subfam.GRUP_CD=fam.GRUP_CD and
subfam.SET_CD=fam.SET_CD and subfam.REP_CD=fam.REP_CD
join FIN_DATA.ANA_GRUP grup on subfam.GRUP_CD=grup.GRUP_CD
and subfam.SET_CD=grup.SET_CD and
subfam.REP_CD=grup.REP_CD
join FIN_DATA.ANA_SET sett on subfam.SET_CD=sett.SET_CD
and subfam.REP_CD=sett.REP_CD
join FIN_DATA.ANA_REP rep on subfam.REP_CD=rep.REP_CD
order by REP_CD;
STORE StrutturaMerceologica INTO [Lib://QVD_folder
(sense_digital_innovation1)/Qvd_StrutturaMerceologica.qvd](qvd
);
```

A.8 Facts Table Accounting Stock

```
[StockContabile]:
LOAD
    PROD_ID,
// PERIOD_ID,
    ENTE_ID,
```

STOCK_QTA_PZ as StockQtaPz, STOCK_QTA_PS as StockQtaPs, DIFF_INVEN_PZ as DiffInvenPz, DIFF_INVEN_PS as DiffInvenPs, //TOT_VEN_VAL as TotVenVal, //TOT_VEN_QTA_PZ as TotVenQtaPz, //TOT_VEN_QTA_PS as TotVenQtaPs, //TOT_VEN_NET_VAL as TotVenNetVal, TOT_COSTO_VAL as TotCostVal, //PROMO_VEN_VAL as PromoVenVal, //PROMO_VEN_QTA_PZ as PromoVenQtaPz, //PROMO_VEN_QTA_PS as PromoVenQtaPs, //PROMO_VEN_NET_VAL as PromoVenNetVal, PROMO_COSTO_VAL as PromoCostVal, MESE_ID, PRZ_MEDIO_ACQ as PrzMedioAcq, VAL_MEDIO_STOCK as ValMedioStock;

SELECT

stock.PROD_ID, stock.PERIOD_ID, stock.ENTE_ID, stock.STOCK_QTA_PZ, stock.STOCK_QTA_PS, stock.DIFF_INVEN_PZ, stock.DIFF_INVEN_PS, 11 stock.TOT_VEN_VAL, 11 stock.TOT_VEN_QTA_PZ, 11 stock.TOT_VEN_QTA_PS, 11 stock.TOT_VEN_NET_VAL, stock.TOT_COSTO_VAL, 11 stock.PROMO_VEN_VAL, 11 stock.PROMO_VEN_QTA_PZ, 11 stock.PROMO_VEN_QTA_PS, 11 stock.PROMO_VEN_NET_VAL, stock.PROMO_COSTO_VAL, stock.MESE_ID,

stock.PRZ_MEDIO_ACQ, stock.VAL_MEDIO_STOCK FROM FIN_DATA.FACT_STOCK stock where PERIOD_ID >=\$(vStartDate); STORE StockContabile INTO [Lib://QVD_folder (sense_digital_innovation1)/Qvd_StockContabile.qvd](qvd);
Bibliography

- [1] ABAX BI METRIX COMPANY, *Reporting*. http://www.bi-metrix.com/isvabak.php.
- [2] ANDREW ABELA, Visual storytelling with data and evidence. http://extremepresentation.typepad.com/blog/2006/09/choosing_{ag}ood.html.
- [3] B. DONNELLAN, C. SHERIDAN E. CURRY, *IT Professional*, A Capability Maturity Framework for Sustainable Information and Communication Technology, 2011.
- [4] BERNARD MARR, Key performance Indicators, Financial Time Press, 2012.
- [5] J. BRESNICK, How health information exchange models impact data analytics, (2015).
- [6] CARL ANDERSON, Creating a Data-Driven Organization, O'REILLY media, Inc., 2015.
- [7] E. CODD, Derivability, redundancy, and consistency of relations stored in large data banks, (1969).
- [8] I. COORPORATION, Two-way anova in spss statistics, (2012).
- [9] F. CORP, Simulation software, (2012).
- [10] S. O. D, The five why's techniches, (2012).
- [11] DAVID T. BOURGEOIS, Information systems for business and beyond. https://bus206.pressbooks.com/.
- [12] DRUCKER, P. F., COLLINS, J., KOTLER, P., KOUZES, J., RODIN, J., RANGAN, V. K., ET AL, The five most important questions you will ever ask about your organization, 2008.
- [13] B. DYKES, Reporting vs. analysis: What's the difference?, (2010).
- [14] H. GREGORY, A guide to enterprise reporting, (2013).
- [15] F. I. N. I. P. E. R. GROUP, Company overview. https://www.iper.it/azienda-impegni-per-il-cliente.php, 2017.
- [16] HAMMER MICHAEL & CHAMPY JAMES, Reingeniería, Editorial Carvajal S.A, 1994.

- [17] T. H. D. J. G. HARRIS, Competing on Analytics: The New Science of Winning, Hardvard Business School, 2006.
- [18] ISO/IEC, *Iso/iec 9075-1:2016*.
 Information technology, Database languages, SQL, Part 1: Framework (SQL/Framework).
- [19] P. JAIN, To centralize analytics or not, that is the question, (2013).
- [20] J. LEEK, The Elements of Data Analytic Style, 2015.
- [21] S. R. M. PIRACCINI, Architetture d'integrazione, (2006).
- [22] T. MCGUIRE, What you need to make big data work: The pencil, (2012).
- [23] MCKINSEY, Essentials of innovation, (2015).
- [24] NATHAN YAU, A day in the life of americans.
- [25] D. J. POWER, A brief history of decision support systems, (2007, March 10).
- [26] QLIK, Qlik community. https://community.qlik.com/welcome, 2017.
- [27] . S. K. RAYNER, N., Maturity model overview for business intelligence and performance management, (2016).
- [28] K. S. RICK KAZMAN, CLAUS NIELSEN, Understanding patterns for system-ofsystems integration, (2013).
- [29] J. STIKELEATHER, Big's datas human component, (2012).
- [30] J. W. TUKEY, The Examination and Analysis of Residuals, Technometrics, 1963.
- [31] US: DEPARTMENT OF TRANSPORTATION, Data integration glossary.
- [32] W. M. P. VAN DER AALST, K. M. VAN HEE, Workflow Management: Models, methods and systems, Eindhoven University of Technology, 2000.
- [33] A. T. M. R. J. WATSON, H. J., Tdata warehousing stages of growth, (2001).