

# Relations between IT Architecture and Information Quality in DW and BI Environments

Samuel Otero Schmidt<sup>1</sup>, Edmir Parada Vasques Prado<sup>2</sup>, Hiroo Takaoka<sup>3</sup>

<sup>1,3</sup>Faculdade de Economia Administração e Contabilidade – Universidade de São Paulo (USP) – CEP 05508-010 – São Paulo – SP – Brazil

<sup>2</sup>Escola de Artes Ciências e Humanidades – Universidade de São Paulo (USP) CEP 03828-000 – São Paulo – SP – Brazil

{schmidt-samuel, eprado, takaoka}@usp.br

**Abstract.** *The organizations are investing more in information technology to collect, store and process vast amounts of information. Generally this information has no standard and repeated data, which hinders the decision making process. The cause of the difficulties is related to Information Quality (IQ) that has technical characteristics related to the architecture used in Data Warehouse (DW) and Business Intelligence (BI) environments. Based on the existing literature and insights from the author a research model was created to identify the relations between components of DW/BI architecture and IQ dimensions. We used the case study as research method. This research makes contributions to better understand IT architecture for DW/BI.*

## 1. Introduction

Investments in information technology (IT) to collect, store and process vast amounts of information are increasing in the organizations. These efforts attempt to translate information in process improvement, better decisions and competitive advantage. One problem faced by many organizations is how to handle a large volume of information. Generally this information has no standard and repeated data, which hinders the decision making process, especially at the strategic level. One of the causes of this difficulty is related to Information Quality (IQ) that has technical features, such as information integration and data warehouse, and non-technical features, such as the lack of strategic cohesion across the organization, so that stakeholders have the right information in the right format at the right time [Madnick et al. 2009]. The technical characteristics of IQ are essentially related to the architecture used in Data Warehouse (DW) and Business Intelligence (BI) environments [Jarke and Vassiliou 1997]. The environment of DW/BI is characterized by its analytical objective, i.e. has the objective of providing IQ for decision makers (senior management, analysts, experts).

The academic literature has many studies about the concept of DW and BI, but few of them are related to the choice of the DW/BI architecture. Some of the early best-known studies on these topics have been researched and presented by Inmon (1996) and Kimball (1996) in the 90's, when many companies began building DW/BI environments [Wixom and Watson 2001]. Both authors helped in the evolution of the analytic environment concept and in the differentiation between this type of environment and the Transactional System (TS) environment.

To Inmon (2005), DW is a more efficient model to store corporate information. This is because, through a centralized database it is possible to share all the information

needed for different business areas, making the DW integrated with historical information - usually with a horizon of five years or more, and with non-volatile information. Kimball and Ross (2002) presents a different architecture with a proposal to quickly meet users' needs through distributed database, known as Data Marts (DM), and through the use of specific information for each business area.

These two authors motivated the discussion of the importance of IT architecture to provide information to businesses, how to obtain proper infrastructure, database, model and IT team to be in an analytical environment, in order to provide users the necessary information so they can make accurate decisions. Further architectures, which are currently used, will be presented on the topic 2.3, adding more components to attend organizations' needs. In this context, one can highlight the importance of IQ to an analytic environment. For Wixom and Watson (2001) the IQ can provide several benefits to users: better understanding of the context of the decision, improves decision making, and changes the way people work. In DW/BI perspective, through appropriate tools of access and searching of information, users are able to make decisions quicker and wide-ranging. In general, DW/BI may change data access by the final users and reduce the required time and effort to provide access to the information [Graham 1996].

In the literature there are various definitions and research on IQ. Authors such as Jarke and Vassiliou (1997), Wang et al. (1995), Wand and Wang (1996), Pipino et al. (2002) present various dimensions to measure the subjective of IQ. In the context of DW/BI, according to Ariyachandra and Watson (2010), the choice of the architecture is an important decision. Studies developed by Meta Group and Gartner identified that architecture selection is one of the aspects that strongly influences the success of a DW/BI solution. A project, named DWQ – Foundations of Data Warehouse Quality [Jarke and Vassiliou 1997; Jarke et al. 1999], created by a research group from some European universities by the end of the 90s, tried to establish a framework, with the intention to analyze DW/BI architecture and IQ factors involved in this perspective. This study produced a wide-ranging framework, presenting architecture with conceptual, logical and physical components and its connections to the IQ dimensions. Despite these efforts, there is still a gap in the literature to spot the relations between the different types of DW/BI architecture, their components and associations with IQ dimensions.

The all-purpose goal of this article is to analyze the components of IT architecture and IQ environments in DW/BI contexts. In order to achieve this general objective, three specific aims were defined: (1) Describe the components of DW/BI environment; (2) Describe the most relevant IQ dimensions in this context; (3) Identify the relations between DW/BI components and IQ in the same environment.

## **2. Literature Review**

### **2.1. Information Quality**

Although there is no consensus on the distinction between Data Quality (DQ) and IQ, there is a tendency to use DQ to refer to technical problems and IQ for non-technical issues. DQ is the syntax of the data, i.e., the structure and form (concerns about the data itself). IQ is about semantics, i.e., the meaning of these data in time, space and context, i.e., to transform information into any component that affects the interpretation and

transformation of data into information [Madnick et al. 2009]. In this paper, this distinction is not made and we adopt the term IQ.

In the literature it is possible to find a range of studies about IQ and there is a considerable attention on the researches that analyze this factor in the DW/BI environment [Wixom and Watson 2001; Arouck 2011]. Based on the literature review with the mainly studies about IQ, some of IQ dimensions were selected to be part of this study and can be considered as dimensions related to DW/BI context presented at the Table 1. These 13 dimensions were selected based on the relation with DW/BI environment presented by the authors in the literature and insights from the authors of this study.

**Table 1. Information Quality Dimensions in DW/BI**

Authors	Dimensions	Definition
Wang et al. 1995; Jarke and Vassiliou, 1997; Wixom and Watson 2001; Michnik et al. 2009; Arouck 2011	Accuracy	The recorded value is in conformity with the actual value
Delone and Mclean 1992; Wand and Wang, 1996; Jarke and Vassiliou, 1997; Pipino et al. 2002	Timeliness	The extent to which the data is sufficiently up-to-date for the task at hand
Wand and Wang 1996; Jarke and Vassiliou, 1997; Wixom and Watson 2001; Pipino et al. 2002	Completeness	The extent to which data is not missing and is of sufficient breadth and depth for the task at hand
Wang et al. 1995; Wand and Wang 1996; Pipino et al. 2002	Reliability	The extent to which data is correct and reliable
Jarke and Vassiliou, 1997; Wixom and Watson 2001; Pipino et al. 2002; Arouck 2011	Consistency	The extent to which data is presented in the same format
Wand and Wang 1996; Jarke and Vassiliou, 1997; Pipino et al. 2002; Michnik et al. 2009; Arouck 2011	Accessibility	The extent to which data is available, or easily and quickly retrievable
Wang et al. 1995; Wand and Wang 1996; Jarke and Vassiliou, 1997; Pipino et al. 2002; Arouck 2011	Interpretability	The extent to which data is in appropriate languages, symbols and units, and the definitions are clear
Wang et al. 1995; Wand and Wang 1996; Jarke and Vassiliou, 1997; Pipino et al. 2002	Believability	The extent to which data is regarded as true and credible
Delone and Mclean 1992; Pipino et al. 2002; Arouck 2011	Conciseness	The extent to which data is compactly represented
Pipino et al. 2002	Quantity	The extent to which the volume of data is appropriate for the task at hand
Wand and Wang 1996; Pipino et al. 2002; Michnik et al. 2009; Arouck 2011	Security	The extent to which access to data is restricted appropriately to maintain its security
Wang et al. 1995; Jarke and Vassiliou, 1997	Availability	The data exists in some form that can be accessed
Pipino et al. 2002	Reputation	The extent to which data is highly regarded in terms of its source or content

## 2.2. Data Warehouse and Business Intelligence Architectures

IT architecture is still a term without a single definition. According to Ross (2003) *apud* Hsing and Souza (2010), “IT architecture is the organization logics to applications, data and concrete technology in a whole of politics and technical choices, which intend to make the company’s business strategy possible.” On the other hand, Spewak and Steven and Steven (1992) states that the technological architecture is the definition for the types of technologies or bases, which will support the business in a data share environment. As the aim of this paper is analyzing the technical aspects of IQ, the definition adopted is the one presented by Spewak and Steven (1992), since the technological architecture is the most adequate term to be related to the DW/BI context.

According to Ariyachandra and Watson (2010), considering that companies spend a million dollars on DW/BI architectures; the lack of academic researches about this subject is surprising. Studies made by Jarke and Vassiliou (1997) highlight the importance of the choice of IT architecture to the success of DW/BI solutions. On Ariyachandra and Watson (2005), Sen and Sinha (2007), Sen and Sinha (2007a), Simoes (2010)’s research, there were stated seven types of architecture to DW/BI: Independent DMs; DM Bus Architecture; Enterprise DW; DW with Operational Data Store (ODS); Hub and Spoke; Federated; and Distributed DW Architecture.

### 2.3. Components of the Data Warehouse and Business Intelligence Architecture

All these architectures have their respective components and when connected they establish the environment architecture. In this study the intention is to evaluate how each component may, positively or negatively affect IQ in DW/BI environment, since the architecture and its components might limit or add to IQ known by the final users. The components considered in this study are:

**Transactional System (TS):** Legacy systems or actual TS execute daily transactions inside the organization, providing data source to DW/BI. e.g: ERP, SCM, data from the market [Simoes 2010; Inmon 2005; Kimball and Ross 2002].

**Staging Area (SA):** It is database or storage area that provides the connection between TS and DW, ODS or DM. According to Kimball and Caserta (2004), depending on the definition of Extract, Transform and Load (ETL), it is possible to use a process in memory. The decision of whether this area will exist or not and how it will be structured basically depend on two goals: (1) Obtain data from TS for the last destiny by the shortest time as possible; (2) Being able to recover data extraction without having to restart the process from the beginning.

**Extract, Transform and Load (ETL):** ETL is made to extract data from TS to other components of analytical environment (SA, DW, DM, ODS). After that, there is transformation, cleaning, integration and treatment of data, in order to load it. The last step is the Load, which is the second part of the integration, updating the information in DW, DM, ODS or MR. According to Sen and Sinha (2007), the extractions and loads usually occurs during batch windows hat process information to the analytical environment during a certain period, in which TS environment or analytical is not available to make transactions or users' access [Cappiello et al. 2008].

**Data Warehouse (DW):** DW is a database, generally relational and normalized, its aiming is to be integrated, time-variant, non-volatile and to share information between all the departments of an organization, in a scalable way a [Kimball and Ross 2002]. In general, it is in DW, where data exists in the lowest level of atomicity [Inmon 2005].

**Data Mart (DM):** DM is a database, in which specific information of departmental areas (e.g: Marketing, Sales, HR, Finance) [Silvers 2008]. In general, DM has aggregated data, based on a dimensional modeling, however, using a DM architecture in bus, there might be an integration of available data through conformed dimensions, what would transform DM on DW [Kimball and Ross 2002]. Regarding this study, it will be adopted the most general definition for DM.

**Operational Data Store (ODS):** ODS is frequently updated and based on what was presented before in this paper, it is made to provide, in an integrated way, operational data extracted from TSs at real time, since the focus of the information is to generate management reports [Silvers 2008; Kimball and Ross 2002].

**Metadata Repository (MR):** MR storages data about data, helping users to navigate through analysis possibilities. It shows how data has to be interpreted by the user, what is the meaning of data, its origin, what transformations it has been through, which are the rules for this transformation, security privilege, quality metrics (i.e: how many times data was accessed). MR can be locally distributed together with DM/ODSs, or centralized with DW [Inmon 2005; Kimball and Ross 2002].

**Data Access (DA):** DA can be considered the last layer to explore and analyze data. In DA, On-line Analytical Processing (OLAP) tools are used to offer a view of the dimensions and facts. [Kimball and Ross 2002].

**Transportation Agent (TA):** TAs are schedules/control program of transferences and network infrastructure, that allow the communication between all components. It is an important component, since DW/BI uses a great volume of data and TA capacities/velocity may be limiting factors to the success of the solution.

### 3. Research Model

The Table 2 presents the relations between the components of DW/BI and IQ dimensions. The impacts of the components in IQ dimensions were defined in a scale with four points: high positive impact, some positive impact, high negative impact and some negative impact. The impacts were interpreted in a qualitative perspective through the information searched in the literature and insights of the author. This paper endeavored to present the main relations found and interpreted as positives/negatives and was not based on an exhaustive research to cover all the possible relations. Through the connections found in this study, it aims to help and motivate future researches on these topics in decision making inside organizations.

**Table 2. Relations between Components and Dimensions**

Key		Dimensions	Components								
TS	Transactional Systems		TS	SA	ODS	ETL	DW	DM	MR	DA	TA
SA	Staging Area	Accuracy	-		+	++	+	-	++		
ODS	Operational Data Store	Timeliness	-	+	++	+	-	+			+/-
ETL	Extract, Transform and Load	Completeness	-			++	++			-	
DW	Data Warehouse	Reliability	++/-		+	++	+	+	++	++	
DM	Data Mart	Consistency	+/-		+	++	++	-	++/-		
MR	Metadata Repository	Accessibility	-	++	+	+	+	++	++		
DA	Data Access Tools (OLAP)	Interpretability				+	++	++	++	+	
TA	Transportation Agent	Believability	++				++	-	++		
++	High Positive Impact	Conciseness	--			+	-	++			
+	Some Positive Impact	Quantity	-			++	++	-		-	-
-	Some Negative Impact	Security		+		++			++		+
--	High Negative Impact	Availability	--	++	+	+	+	+		+	-
		Reputation	++/-						++	++	

**Accuracy:** TS is one of the components that most impacts data precision, since data that is inserted in this layer will permanently influence all the other steps and a mistake at this level cannot be corrected, for the reason that once the fact occurred it is not possible to repeat it, the majority of the these inaccuracies happens during the insertion in this stage. Both ODS and DW positively impact on this dimension, because, in general, information stored in the database was already treated by ETL, what avoid syntax mistakes and redundancy, improving the accuracy. On the other hand, DM may have a negative impact due to the fact that information might be spread in various DMs and also be extracted in different sources. ETL aims to treat and clean data while MR presents the real meaning for data and its standard, avoiding conformation mistakes.

**Timeliness:** TS has a negative impact on this dimension, since the transactional cannot be available or be accessed during high utilization, so as the information can be extracted to the analytical environment. On the other side, SA has a positive impact, because if there is any problem related to the ETL processing and if it is necessary to restart the process, this component avoids searching the data again in ST. ODS is made

to have the information in a more immediate way when compared to DW or DM, that why is has a high positive impact on this dimension [Silvers 2008]. DM can be directly updated, avoiding keeping the information that will be loaded in DW and also after that. TA influences the time for transferring a great volume of data.

**Completeness:** This dimension is negatively influenced by ST, in which profundity and amplitude of information are modeled. ETL may improve completeness through the data extraction from different sources, in various levels of granularity and integrate this information in DW, where it is possible to allocate a great volume of data in the lowest level of granularity. Further on, CE may have a negative impact on this dimension, because depending on the query or cube created and executed by the user it is possible to get not incomplete information.

**Reliability:** TS might impact on both positive and negative manners, because, in case users' community realizes that the source of data is reliable, TS will have a positive impact, however if the client realizes that it is not, TS will be avoided. ETL has an important role to evaluate reliability, because, based on business rules in transformation process, one can verify how much data in reliable and besides, it is possible to load a reliability metric of data inside MR. MR may serve as a repository of quality metrics, so that user is able to evaluate the quality of decision, regarding obtained IQ. In CE is where reliability is approved, since users are most interested on validating information and through the analysis they will identify if data is reliable or not.

**Consistency:** TS may have a positive or negative impact on data consistency. The negative impact occurs when data is inserted twice. In this case, this information will have treated by ETL, which has a fundamental role to guarantee the consistency, since DW should not be loaded twice with the same information, in this way, one can try to obtain a single version of information. Both DW and ODS have a positive impact on this dimension, because they aim to receive integrated data and consequently enhance the consistency. DM might have a negative impact, for the reason that the same data might be loaded in different DMs and frequently, there is not a single version of information. MR might have a positive impact in case there is an single and centralized MR and through this components, it is feasible to analyze how many times that information was loaded in DW. DM or ODS. Nevertheless, in case there are many local MRs for each DM, there might be a high negative impact on this dimension.

**Accessibility:** ETL is made to load data in ODS, DW and DM with access permission for users and make the access of data possible in the different types of components. MR can also receive access information e control this in its repository, in this way, it is likely to keep a history of information access. SA positively influences this dimension, since it can recovers data without the necessity of accessing it again in ST. DM enhances the accessibility, for the reason that there are specific information of each department, what makes the information closer to the final user, improving access control and proving an easy and quick recovery.

**Interpretability:** MR has a high positive impact on this dimension, due to its contribution presenting the meaning of the introduced information, how it should be used and also what is the context related to the actual fact. MR should be strongly integrated to CE, in order to make the significance of information understandable to users. ETL influences interpretability, because it establishes the business rules treat information. DW may enhance interpretability, because this component contains the

lowest level of granularity, in which it is possible to understand the information, in a detailed way, and make an accurate interpretation about it. DM also supports this dimension, for the reason that data will be closer to the subjects related to the users.

**Believability:** One can highlight the positive impact of ST, because it provides the origin of data what enables comparison in order to validate it. It is also possible to analyze whether the source of data is adequate or not. According to Inmon (2005), DW may positively influence credibility, because users can receive integrated information. On the other side, DM might have a negative impact on this dimension, due to extractions from different sources. In this case, MD's role is to inform the origin of information, which steps it, has been through that represents how much this data is believable [Strong et al. 1997].

**Conciseness:** One can highlight that DM has a high positive impact on this dimension, because in this component there are aggregated and summarized data, what compacts information. Besides, DW, as stated before, has a great volume of data with low levels of granularity, what would have a negative impact on this dimension. ETL aims to compact data, through the cleaning. In TS there are no compacted data in analytical environment, so it may have a negative impact.

**Quantity:** TS has a negative impact on data quantity, because, generally, there is no window for a long period to the extraction of data through batch routines. ETL aims to bring enough quality to the analytical environment in order to attempt the requirements of information from users. DW positively influences this dimension, because this component has a great volume of data that may be accessed by users, however, DM is limited in terms of data quantity. CE and TA may have a negative impact on these dimensions, since in CE, user may search for an unnecessary data quantity or bring a great volume of data that may compromise TA capacity and in this case, both would have a negative impact.

**Security:** SA has high positive impact on this dimension, because at this layer it is possible to limit the access of other components. In general, only ETL, has access to SA. ETL plays a key role in this dimension, because it can load data with safety privileges and, besides, it also may mask data so as users are able to analyze it without accessing confidential information. The privilege control is generally stored in MR. TA may positively impact, because, usually, this analytical environment has a segregated infrastructure from TS infrastructure environment, avoiding safety risks.

**Availability:** TA may generate a negative impact on the access to the information, in case the network is not available. In TS case, there is a high negative impact, because the focus of TS availability is to its operational needs and do not seek relation to information, that may be available to the analytical environment. In general, there are specific windows to the availability of data and many STs need to be offline, so as ETL Extraction is made through batch routines [Strong et al. 1997]. If the choice is to directly access ST, there may be a competition with the operational processing and access of this component. SA aims to minimize TS impact, since its availability is focused on analytical environment and allows the recovery of information without accessing TS again. ODS, DW, DM and CE can also have positive impacts, because their availabilities are hearted on analytical environment users as well.

**Reputation:** TS has the same impact introduced to reliability. When data is underutilized by users, MR may keep this use metric, that shows users whether that

information has a good reputation or not. Besides, RMS should improve data interpretation that consequently, may enhance the reputation. In this case, CE has the same impact identified to reliability.

#### **4. Research Method**

This research is characterized as a qualitative study. According to Selltiz and Deutsch (1975), qualitative research is suitable to describe the complexity of a given situation and understand their dynamic processes. As a result, this type of research is appropriate for this work, because it analyzes the IT architecture and IQ in DW/BI environments. Additionally, one can classify this research as exploratory. According to [Selltiz and Deutsch, 1975], the exploratory research serves to increase the knowledge of the researcher about the phenomenon under investigation, and serves as the basis for the formulation of problems to future research.

##### **4.1. Research Strategy**

The case study is a method of empirical research that investigates contemporary phenomena in their real context, where the boundaries between phenomenon and context are not clearly defined [Yin 2005]. Moreover, it is especially useful in the generation and theory building, where little data or theory exist [Eisenhardt 1989]. The research strategy adopted in this paper was the study of a single case, because the focus was analyze and validate in deep the impact of architecture and components in DW/BI in IQ dimensions as an exploratory research. So was necessary to select a big company that has high investments in DW/BI environment.

##### **4.2. Object of study**

The case study was conducted at a big Financial Company (FC) in Brazil. The real name of the company will not be presented, so in this paper the company will be called as FC. FC has more than 30 thousands employees in total and more than two thousands at IT Area. The main products of this company are account transactions, credit cards, internet banking, and insurance. There are offices and agencies spread out in all states in Brazil, and other countries as Argentina, Chile, and European Continent. The revenue is more than R\$ 30 billion per year and expends more than R\$ 1 billion in DW/BI per year to maintain and implement new projects. There area more than three thousands users for DW/BI systems using Data Stage and PowerCenter (ETL tools); Statistical Analysis System (SAS), as ODS database; DB2 and Oracle (as DW and DM); MicroStrategy and Cognos (for OLAP tools). The most part of the source of the information comes from the Mainframe (the TS). The type of the architecture currently used is a Distributed DW, because they have distributed ETL processing running for the legacy DMs, and now they are implementing a new architecture based on Hub and Spoke.

##### **4.3. Data Collection**

The instrument used for data collection was an interview. An original script of questions was developed and applied individually to each one of the interviewees. The interviews were conducted in the second half of 2011 with five people who work at IT department and had experience in DW and BI (At the Table 3).

**Table 3. Interviewees Profiles**



IT DW/BI Areas	Position	Years of experience in DW/BI environments	Years working in the company
ETL Team	Cordinator	8	26
ETL Team	Senior Analyst	7	1
ETL Team	Senior Analyst	12	2
Support Infrastructure	Senior Analyst	6	23
Data Access Tools	Junior Analyst	3	1

It is possible to verify that most of the respondents have a lot of years of experience in DW/BI environments, so it was easy to interview them about the architecture and components of DW/BI, but this was not the case in relation to IQ dimensions. In this case was necessary to explain some concepts to clarify what is IQ in this study and literature. After that, the interviewees understood the IQ concept adopted in this study, and they easily could associate this concept with their daily activities.

## 5. Data Analysis and Results

In this study was possible to verify some associations between IQ dimensions and DW/BI Architecture. At first, at the FC was possible to get how the IT DW/BI team sees the importance of IQ dimensions and DW/BI components. At the Table 4 shows a ranking of importance for IQ dimensions which is a summary of the data collected at the interviews. The results show that Accuracy, Reliability and Believability were more important than others, and it makes sense and according with [Sen and Sinha 2007; Ariyachandra and Watson 2005; Jarke and Vassiliou 1997]. The accuracy is related with the daily users' activities, because they need accurate information to send to Government, or to investors and stakeholders. The consistency is related with to necessity to have a single version of the truth, and the security is really important for FC, because there is a high risk of violation data by the employees, or crimes by hackers. Availability is the main objective of the structure responsible for the infrastructure support (presented at the topic 4.2). So, after providing quality for users' necessities, the mission of the IT area is to offer availability at the infrastructure level. Reputation is related to the importance to understand where is the source of the data (e.g. TS) and how is IQ in this source. At FC the new ETL is trying to provide a better MR and integration to show which are the sources of all information in DW/BI. For IT analysts the completeness is improved by the integration at the ETL level, but they don't know clearly how the user see this dimension, so they don't have a perception of high importance to it. Accessibility is important, because the daily operations is too critical that runs in the TS (or *Mainframe*), and cannot be accessed any time during the day, just in *batch* window. This is important to avoid high concurrence in the TS processes to not impact operations. Actually the timeliness is not so critical, the frequency of Load usually is daily, and so the users have a one day delayed view. The necessity of quantity is common for DW/BI systems specially using DW components. Interpretability is also important and it is intrinsic with BI applications, e.g: MR component. Conciseness is also common, but actually the user just thinks about the necessity of information and not about the conciseness. Some dimensions were identified as more important than others, and is possible to verify the intrinsic differences in the necessities of IQ in a DW/BI environment at the FC.

**Table 4. Dimensions Ranked by Importance in DW/BI Environment**

Dimensions	Importance Ranking
Accuracy, Reliability, Believability	1
Consistency, Security	2
Availability	3
Completeness, Reputation	4
Accessibility	5
Timeliness, Interpretability, Quantity	6
Conciseness	7

At the Table 5 is possible to identify the most important components in the DW/BI environment. For the interviewees' perception, the ETL, DW and DM are the most important components for the new structure. It's consistent with the research model (at Table 2).

**Table 5. Components Ranked by Importance in DW/BI Environment**

Components	Importance Ranking
ETL, DW, DM	1
ODS, MR, DA, TA	2
SA	3
TS	4

The ODS is also really important, but it is not so structured as an ODS, sometimes the SAS applications work as DMs, and other times as an ODS. For the new architecture, FC wants to create a new ODS for SAS users with a high quality of timeliness, consistency, and security. Actually each department has its own SAS data sets, and for the new structure it is important to integrate all this date sets in a single ODS component. Actually they see the importance of MR, and they want to improve the use of this component centralizing it in a single a MR. At now there are distributed MR that worse the consistency. DA is really important to provide different views of the information and good interpretability using cubes and graphics. TA is necessary to transport a huge volume of data, and maintain the security. In the forth place the TS is not viewed as an important component, it is just necessary to extract the data. It is also consistent with research model, because usually the TS is not implement for DW/BI proposes, instead its focus is on operational requirements.

## 6. Conclusions and Recommendations

This qualitative study examined the relations between DW/BI architectures and IQ dimensions by using an empirical case of study, in a single company from financial market in Brazil, based on a research model that was developed using DW literature, IQ literature and insights from author. This paper presents two preliminary contributions:

**(1) There are IQ dimensions that are more important in DW/BI environments:** In this literature there are a lot IQ dimensions, and in this study were selected 13 dimensions (Table 1), that were more founded in the literature of IQ and DW/BI architecture [Wixom and Watson 2001; Cappiello et al. 2008]. The case of study at FC showed that there is a ranking for perception of importance for some IQ dimensions. This find is important for the organizations, because it recommends they identify and prioritize some IQ dimensions to guide and improve the success of the DW/BI architecture. It was not found in the literature a study identifying the more important dimensions for DW/BI context.

**(2) The components of DW/BI architecture are associated with IQ dimensions:** Based on the literature review and insights was possible to create a research model with

the relations between components of DW/BI architecture and IQ dimensions (at the Table 1), and using the case of study was possible to verify some relations are true in the reality of FC. At the research model were identified negatives and positives impacts in all dimensions. The ETL, DW and DM were the components that most generated positives impacts on dimensions. It happened, because all three components were developed specifically for the necessity of information requested by users to make better decisions [Inmon 2005; Silvers 2008; Kimball and Ross 2002]. According to Ariyachandra and Watson (2010), and Silvers (2008), and the case of study at FC, in fact the all three components have extreme importance in DW/BI. On the other hand, it is interesting to analyze the high quantity of negatives impacts on TS and TA. The negatives impacts on TS could be interpreted as this component usually is not implemented for analytical purpose, therefore its main objective is to support operational requirements. It is a valuable contribution, because actually the companies to get more competitive should plan and to implement the TS not just for operational environment requirements, but also for analytical environment requirements. The TA usually has negatives impacts, because the users want to get huge volume of data in less time possible and, many times, the TA is limited in throughput to transport it.

There are some limitations in this research: It is important to note that this is preliminary study. It was not possible to cover all the potential relations between components of DW/BI and IQ dimensions that could exists in the literature. It is necessary more literature review; and using study of case is not possible to generalize the results. Based on the considerations presented in this paper, it is recommended to elaborate future works: Replicate the interviews with DW/BI users to understand better how the components could be related with IQ; Make new research to identify which dimensions are important for DW/BI; Make new studies analyzing the impact on the different types of DW/BI architecture in IQ.

## References

- Ariyachandra, T. and Watson, H. (2005). *Key Factors in Selecting a DW Architecture* BI Journal; Spring.
- Ariyachandra, T. and Watson, H. (2010). Key organizational factors in DW architecture selection, *Decision Support Systems (DSS)*, v.49, n.2, p.200-212, May.
- Arouck, O. (2011). *Atributos de qualidade da informação*. Dissertação – UNB, Brasília.
- Cappiello, C. et al. (2008) “Enabling Better Decisions through Quality-Aware Reports in BI Applications”. In *Proceedings of the 13th ICIQ*, MIT Press, Boston, MA, USA.
- DeLone, W.H. and McLean, E.R. (1992). "Information Systems Success: The Quest for the Dependent Variable," *Information Systems Research* (3:1), pp 60-95,
- Eisenhardt K. M. (1989). *Building Theories from Case Study Research*. *Academy of Management Review*, v. 14, n 4, p. 532-550.
- Graham, S. (1996). *The Foundations of Wisdom: A Study of the Financial Impact of DW*, International Data Corporation, Toronto.
- Hsing, C. W. and Souza, C. A. (2011). *Influências Institucionais e Estratégicas em Decisões de Arquitetura de TI: Estudo de Casos Comparados em Empresas Brasileiras*. In: *ENADI 2011 - III Encontro de Administração da Informação*.

- Inmon, W.H. (1996). Building the DW, 2<sup>nd</sup> ed, John Wiley & Sons Inc. New York.
- Inmon, W. H. (2005). Building the DW. 4<sup>th</sup> ed, Wiley, New York.
- Jarke, M. and Vassiliou, Y. (1997). Foundations of Data Warehouse Quality. Esprit.
- Jarke, M., Jeusfeld, M. A., Quix, C. And Vassiliadis, P. (1999). Architecture and Quality in DW: An Extended Repository Approach. Inf. Syst. 24(3): 229-253.
- Kimball, R. (1996). The DW Toolkit, John Wiley & Sons Inc., New York.
- Kimball, R. and Ross, M. (2002). The DW Toolkit: The Complete Guide to Dimensional Modeling, 2nd edition, Wiley, New York.
- Kimball, R. and Caserta, J. (2004). The DW ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. Wiley Publishing Inc.
- Madnick, S. E et al. (2009) Overview and Frame work for Data and IQ Research, ACM Journal of Data and IQ, v.1, n.1, Article 2.
- Moss, L. T and Shaku, A. (2003) BI Roadmap: The Complete Project Lifecycle for Decision-Support Applications Addison Wesley.
- Pipino, L., Lee, Y., and Wang, R. (2002). Data Quality Assessment. Communications of the ACM (CACM), 45(4). New York, NY, USA.
- Ross, J. W. (2003). Creating a Strategic IT Architecture Competency: Learning in Stages. MIS Quarterly Executive.
- Selltiz; J. and Deutsch, C. (1975). Métodos de Pesquisa nas Relações Sociais. São Paulo: Editora da Universidade de São Paulo.
- Sen, A. and Sinha, A. P. (2007). Toward developing DW process standards: an ontology-based review of existing methodologies, IEEE, Man and Cybernetics 37 (1).
- Sen, A. and Sinha, A. P. (2007a). A comparison of DW methodologies. CACM, p.79–84.
- Simoës, D. (2010). Conceptual Framework for the Construction and Evaluation of DW, ,UA, Aveiro, Portugal.
- Silvers, F. (2008) Building and Maintaining a DW, Auerbach.
- Spewak, K. and Steven H. (1992). ERP, Developing a Blueprint for Data, Applications and Technology. NY: John Wiley & Sons.
- Strong, D. M. and Lee, Y. W. (1997). Data Quality in Context. CACM, May, v.40
- Wand, Y. and Wang, R. Y. (1996). Anchoring DQ dimensions in ontological foundations. CACM, v. 39, n. 11, p. 86-95.
- Wang, R. Y. et al. (1998). Manage your information as a product. Sloan Management Review, v. 39, n. 4, p. 95-106, Summer.
- Wang, R. Y. et al. (1995). Toward DQ: An attribute-based approach. DSS, v.13, p.349-372.
- Wixom, B. B. H. and Watson, H. J. (2001). Research Article an Empirical Investigation of the Affecting DW Success. MIS Quarterly, v. 25, n. 1, p. 17-41.
- Yin, R. K. (2005). Estudo de caso planejamento e métodos. Porto Alegre: Bookman.