



**UNIFACS**

UNIVERSIDADE SALVADOR

LAUREATE INTERNATIONAL UNIVERSITIES\*

**PROGRAMA DE PÓS-GRADUAÇÃO  
MESTRADO ACADÊMICO EM SISTEMAS E COMPUTAÇÃO**

**MARCIO HENRIQUE ALVES DOS SANTOS**

**UM MODELO PARA A AVALIAÇÃO DA QUALIDADE DE DADOS EM SISTEMAS  
TRANSACIONAIS**

Salvador  
2014

**MARCIO HENRIQUE ALVES DOS SANTOS**

**UM MODELO PARA A AVALIAÇÃO DA QUALIDADE DE DADOS EM SISTEMAS  
TRANSACIONAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Sistemas e Computação da Universidade Salvador – UNIFACS, como requisito parcial para obtenção do grau de Mestre.

Orientador: Prof. Dr. Paulo Caetano da Silva.

Salvador  
2014

## FICHA CATALOGRÁFICA

(Elaborada pelo Sistema de Bibliotecas da UNIFACS Universidade Salvador, Laureate International Universities)

Santos, Marcio Henrique Alves

Um modelo para a avaliação da qualidade de dados em sistemas transacionais. / Marcio Henrique Alves dos Santos. - 2015.

140 f. : il.

Dissertação Programa de Pós-Graduação em Sistemas e Computação de UNIFACS Universidade Salvador, Laureate International Universities como requisito parcial à obtenção do título de Mestre em Sistemas e Computação.

Orientador: Prof. Dr. Paulo Caetano da Silva.

1. Sistema de computação. 2. Qualidade de Dados. 3. Trigger. 4. Sistemas Transacionais. I. da Silva, Paulo Caetano, orient. II. Título.

CDD: 005. 1

MÁRCIO HENRIQUE ALVES DOS SANTOS

UM MODELO PARA A AVALIAÇÃO DA QUALIDADE DE DADOS EM SISTEMAS  
TRANSACIONAIS

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Sistemas e Computação da Universidade Salvador, para obtenção do grau de Mestre em Sistemas e Computação.

Paulo Caetano da Silva Orientador - \_\_\_\_\_  
Doutor pela Universidade de Pernambuco – UFPE  
Universidade Salvador – UNIFACS

Glauco de Figueiredo Carneiro \_\_\_\_\_  
Doutor pela Universidade Federal da Bahia – UFBA / Universidade Salvador - UNIFACS  
Universidade Salvador – UNIFACS

Sidney da Silva Viana \_\_\_\_\_  
Doutor pela Universidade de São Paulo – USP  
Centro Universitário Dinâmica das Cataratas

Salvador, 10 de março 2014.

Ao meu pai  
Saudade, Amor, Sabedoria.  
À minha mãe,  
Fé, Amor e Perseverança.

À Alessandra,  
Amada Esposa e Companheira.

## **AGRADECIMENTOS**

Agradeço primeiramente a Deus, que sempre acompanhou todos os passos de minha vida, não me permitindo fraquejar nos momentos de dificuldade.

À minha linda esposa Alessandra por todo o seu amor, paciência, companheirismo e compreensão durante mais esta jornada.

Ao meu querido pai que não se encontra mais entre nós, mas sempre me apoiou e em todas as minhas decisões.

À minha mãe que com sua sabedoria, amor e dedicação me fez ser o homem que hoje sou.

Ao Professor Dr. Sidney Viana que, com muita paciência, me ajudou a definir o tema deste trabalho.

Ao meu orientador, Professor Dr. Paulo Caetano da Silva, pelo carinho e dedicação despendidos a este trabalho, sem o qual, o seu desenvolvimento não teria sido possível.

## RESUMO

A qualidade dos dados pode ser um fator determinante no desempenho de uma organização. A baixa qualidade dos dados pode impactar negativamente, fazendo com que os processos de tomada de decisão sejam comprometidos. Os problemas de qualidade podem advir das mais diferentes fontes, sejam elas erros de processamento dos software, falhas na determinação de requisitos ou falhas no processo de carga dos dados nos sistemas. Para viabilizar o provimento de informações acerca da qualidade dos dados em sistemas transacionais sob a ótica do usuário, é apresentado nesta dissertação um Modelo para a Avaliação da Qualidade de Dados em Sistemas Transacionais, no qual é proposta uma ferramenta, FEMEQ, para auxiliar a medição da qualidade dos dados de maneira automatizada, através do uso de triggers e um modelo de processo para a avaliação da qualidade de dados sob a perspectiva do usuário. Para aplicação do modelo proposto com base no processo e na FEMEQ foi realizado um estudo de caso em três organizações que utilizam software para domínios distintos de modo que a qualidade de dados nesses sistemas pudesse ser inferida. Este estudo de caso permitiu identificar a validade do modelo proposto, bem como benefícios auferidos pela sua utilização.

**Palavras-chave:** Qualidade de Dados. Trigger. Sistemas Transacionais e TDQM.

## ABSTRACT

Data quality can be a determining factor in the performance of an organization . The poor quality of data can negatively impact , making the decision -making processes are compromised. Quality problems can arise from many different sources , whether the software processing errors , failures in determining requirements or failures in the data load process in the systems . This paper presents a Model for Assessing the Quality of Data in Transactional systems from the perspective of the user. Its function is to facilitate the provision of information about the quality of data in transactional systems. Are presented on a tool, FEMEQ that uses triggers to assist in measuring the quality of data in an automated way and a process model for the assessment of data quality from the perspective of user. For the application of the proposed model based on process and FEMEQ a case study was conducted in three organizations that use software to different domains so that the quality of data in these systems could be inferred. This case study allowed us to identify the validity of the proposed model as well as benefits accrued by its use.

**Keywords:** Quality Data, Trigger, Transactional Systems and TDQM.



## LISTA DE ILUSTRAÇÕES

Figura 1 - Representação da ocorrência de problemas de qualidade.....	16
Figura 2 - Etapas para a Aplicação da TDQM .....	32
Figura 3 - Esquema do Modelo Entidade Relacionamento com as modificações necessárias para medição da Qualidade da SMI e da informação.....	45
Figura 4 – Funcionamento de uma trigger.....	56
Figura 5 - Arquitetura da FEMEQ.....	60
Figura 6 - Modelo Lógico de Dados da Ferramenta.....	64
Figura 7 - Legenda de Símbolos UML.....	66
Figura 8 - Diagrama de Atividades para Avaliação de Qualidade Utilizando a FEMEQ.....	67
Figura 9 - Modelo Lógico do Sistema em avaliação com as alterações realizadas pela FEMEQ .....	72
Figura 10 - Atuação da Trigger na Base de Dados do Sistema Avaliado .....	73
Figura 11 - Determinação das Dimensões de Qualidade de Dados.....	77
Figura 12 - Modelo Lógico do Cliente A .....	86
Figura 13 - Modelo Lógico do Cliente A .....	87
Figura 14 - Modelo Lógico do Cliente C .....	88
Figura 15 - Resultado dos testes de qualidade para a credibilidade no estorno de mensalidade .....	90
Figura 16 - Resultado dos testes de qualidade para a Completeza na venda com o Cartão .....	91
Figura 17 - Resultado dos testes de Segurança no cancelamento das vendas .....	91
Figura 18 - Resultado dos testes de credibilidade no lançamento da telefonia .....	92
Figura 19 - Resultado dos testes de Temporalidade na hospedagem do cliente.....	92
Figura 20 - Resultado dos testes para Completeza no cancelamento de aluguel. ....	93
Figura 21 - Resultado dos testes de Completeza no registro de frequência dos associados.....	94
Figura 22 - Resultado dos testes de Credibilidade no Registro de Entrada de Sócios .....	94
Figura 23 - Resultado dos testes de Temporalidade no lançamento da mensalidade.....	95
Figura 24 - Resultado dos testes para a Quantidade Adequada de Dados no lançamento de boletos.....	95

## LISTAS DE TABELAS

Tabela 1 - Quadro Metodológico.....	22
Tabela 2 - Termos de busca na Web.....	24
Tabela 3 - Dimensões da qualidade.....	35
Tabela 4 - Características para a avaliação de qualidade de dados .....	53
Tabela 5 - Dimensões apuradas para o Cliente A, Cliente B e Cliente C .....	83
Tabela 6 - Relação entre dimensões, valores e métricas utilizados para os testes da QI nos clientes.....	84
Tabela 7 - Resultados obtidos pela FEMEQ .....	89
Tabela 8 - Resultados obtidos durante a avaliação.....	97
Tabela 9 - Características para a avaliação de qualidade de dados .....	99

## LISTA DE ABREVIATURAS E SIGLAS

AIMQ	Assessment and Improvement Methodology for Quality
AQUA	Ferramenta para Avaliação da Qualidade de Dados pela Não Conformidade
AQUAWARE	Ambiente de Suporte à Qualidade de Dados em Data Warehouse
API	Application Programming Interface
CMM	Capability Maturity Model
CRM	Customer Relationship Management
DDL	Data Definition Language
DML	Data Manipulation Language
ERP	Enterprise Resource Planning
FEMEQ	Ferramenta de Medição de Qualidade de Dados em Ambientes Transacionais
GQM	Goal Question Metric
IQA	Information Quality Analysis
ISO	International Organization for Standardization
ISO/IEC	International Standards Organization/International Electrotechnical Commission
OO	Orientado a Objeto
MDI	Multiple Document Interface
ME-R	Modelo Entidade-Relacionamento
MRP	Materials Requirements Planning
MVC	Model-View-Control
PDCA	plan, do, check e act
PL/SQL	Procedural Language/Structured Query Language
QI	Qualidade da Informação
RAD	Rapid Application Development
RMQ	Repositório de Metadados da Qualidade
RUMBA	Reasonable, Understandable, Measurable, Believable e Achievable
SAP/CRM	Aplicação CRM da empresa SAP
SDI	Single Document Interface
SGBD	Sistema Gerenciador de Banco de Dados
SI	Sistema de Informação
SMI	Sistema de Manufatura da Informação
SPICE	Software Process Improvement and Capability Determination

SQL	Structured Query Language
SQL3	Structured Query Language padrão 3
TDQM	Total Data Quality Management
TQM	Total Quality Management
TRAQ	Timeliness + Reliability + Accuracy = Quality
UML	Unified Modeling Language
XMLA	Extended Markup Language for Analysis
XMLA-Q	Extended Markup Language for Analysis of Quality

## SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO.....	15
1.1 CONTEXTUALIZAÇÃO .....	15
1.2 JUSTIFICATIVA .....	17
1.3 MOTIVAÇÃO.....	18
1.3 OBJETIVOS .....	19
1.3.1 Objetivo Geral .....	19
1.3.2 Objetivos Específicos .....	20
1.4 ORGANIZAÇÃO DA DISSERTAÇÃO .....	20
CAPÍTULO 2 - METODOLOGIA .....	22
2.1 MÉTODO ESTRUTURALISTA .....	22
2.2 METODOLOGIA DA PESQUISA.....	23
2.2.1 Pesquisa Bibliográfica .....	23
2.2.2 Pesquisa com Survey .....	25
2.2.3 Pesquisa Experimental.....	25
2.2.4 Natureza Aplicada .....	25
2.2.5 Abordagem Qualitativa.....	26
2.3 CONSIDERAÇÕES FINAIS .....	26
CAPÍTULO 3 - QUALIDADE DE DADOS .....	27
3.1 DADO, INFORMAÇÃO E CONHECIMENTO .....	27
3.2 QUALIDADE DE DADOS.....	28
3.3 TOTAL DATA QUALITY MANAGEMENT - TDQM .....	30
3.4 QUALIDADE DE SOFTWARE.....	33
3.5 DIMENSÕES DA QUALIDADE .....	34
3.6 CONSIDERAÇÕES FINAIS .....	36
CAPÍTULO 4 - ESTADO DA ARTE .....	38
4.1 METODOLOGIAS DESENVOLVIDAS PARA AVALIAÇÃO DA QUALIDADE DOS DADOS .....	38
4.2.1 Avaliação da qualidade pela não conformidade .....	40
4.2.2 Avaliação da qualidade em Data Warehouses.....	42
4.2.3 Avaliação da qualidade em Ambientes Transacionais .....	43
4.3 AVALIAÇÃO DA QUALIDADE SOB A PERSPECTIVA DO USUÁRIO .....	47
4.4 FERRAMENTAS DE AVALIAÇÃO DE QUALIDADE DE DADOS.....	51
4.5 CONSIDERAÇÕES FINAIS .....	52
CAPITULO 5 – UM AMBIENTE PARA A AVALIAÇÃO DA QUALIDADE DE DADOS EM AMBIENTES TRANSACIONAIS .....	54
5.1 FEMEQ - FERRAMENTA PARA MEDIÇÃO DA QUALIDADE EM BANCOS DE DADOS TRANSACIONAIS .....	54

5.1.1	Premissas para o Desenvolvimento da FEMEQ.....	54
5.1.2	Requisitos Funcionais e Não Funcionais.....	57
5.1.3	Arquitetura da FEMEQ .....	59
5.1.4	O Modelo Lógico de Dados da FEMEQ .....	62
5.1.5	Implementação da FEMEQ .....	64
5.2	MODELO DE PROCESSO PARA AVALIAÇÃO DA QUALIDADE DOS DADOS BASEADO NA PERCEPÇÃO DOS USUÁRIOS .....	66
5.2.1	Adição de Entidades de Dimensão ao Modelo Lógico do Sistema em Avaliação.....	70
5.2.2	Monitoramento dos Dados .....	72
5.2.3	Geração dos Relatórios.....	74
5.2.4	Uma Proposta para a Definição das Dimensões de Qualidade sob a Perspectiva do Usuário .....	76
5.3	CONSIDERAÇÕES FINAIS .....	78
	CAPITULO 6 – Estudos de Caso Utilizando a FEMEQ.....	80
6.1	INTRODUÇÃO.....	80
6.2	APRESENTAÇÃO DOS CENÁRIOS.....	80
6.3	A METODOLOGIA DE AVALIAÇÃO.....	81
6.4	ADIÇÃO DAS TABELAS DE DIMENSÕES NO MODELO LÓGICO DE DADOS DOS CLIENTES .....	85
6.4.1	Cliente A.....	85
6.4.2	Cliente B .....	86
6.4.3	Cliente C.....	88
6.5	RESULTADOS OBTIDOS .....	89
6.5.1	Cliente A.....	90
6.5.2	Cliente B .....	91
6.5.3	Cliente C .....	93
6.6	LIMITAÇÕES E RESTRIÇÕES DO ESTUDO DE CASO .....	95
6.7	CONSIDERAÇÕES FINAIS .....	96
	CAPITULO 7 - CONCLUSÕES E TRABALHOS FUTUROS .....	98
7.1	CONCLUSÕES .....	98
7.1.1	Principais Contribuições.....	100
7.1.2	Trabalhos Futuros .....	101
	REFERÊNCIAS .....	103
	APÊNDICE A - DICIONÁRIO DE DADOS .....	107
	APÊNDICE B - TRIGGERS DE TESTES .....	110
	APÊNDICE C - QUESTIONARIO DE ENTREVISTA DE CLIENTE.....	129
	APÊNDICE D - SCRIPTS PARA A CRIAÇÃO DAS TABELAS DE DIMENSÕES.....	133

# CAPÍTULO 1 - INTRODUÇÃO

## 1.1 CONTEXTUALIZAÇÃO

Segundo Pressman (2006) “qualidade de software é a satisfação de requisitos funcionais e de desempenho explicitamente declarados, normas de desenvolvimento explicitamente documentadas e características implícitas que são esperadas em todo o software desenvolvido”, o que, em outras palavras, quer dizer que a qualidade de software está mais intimamente relacionada aos resultados apresentados ao usuário e às técnicas de desenvolvimento do que às tecnologias aplicadas. Entre os fatores que podem impactar diretamente nos resultados apresentados aos usuários dos sistemas, como afirma Wang *et al* (2000), está a qualidade de dados. A observação da qualidade dos dados gerados por um sistema e o seu impacto para o usuário servirão de base, neste trabalho, para a medição da qualidade do sistema.

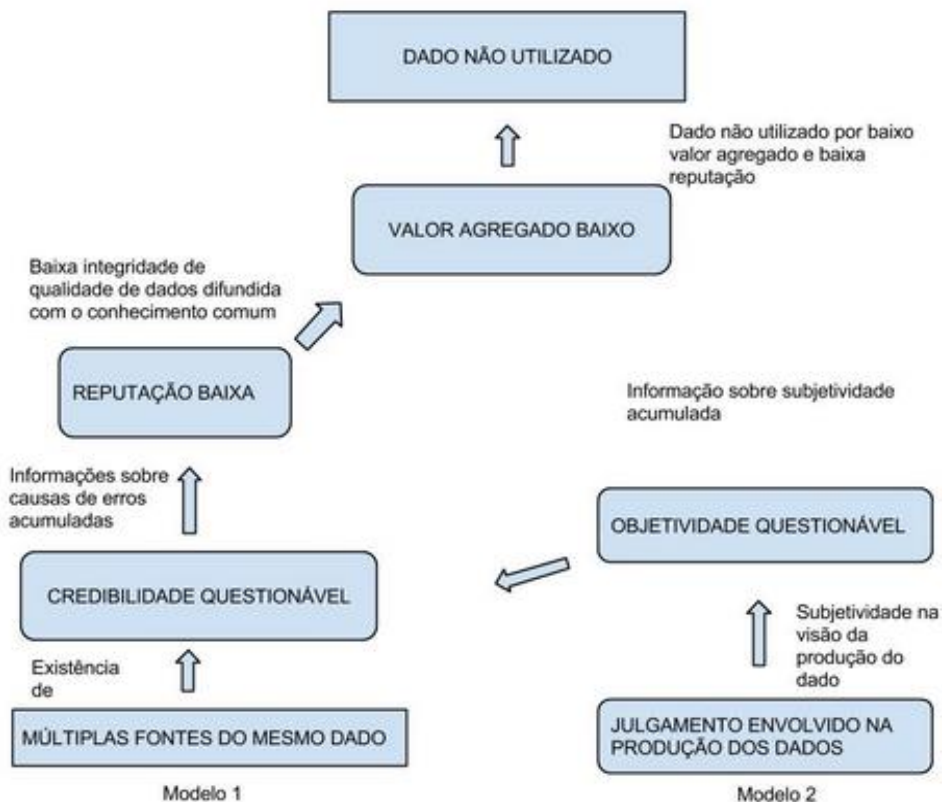
Segundo Moss (2003), as organizações em todo o mundo desejam obter qualidade em seus dados, porém, normalmente a qualidade desejada não é encontrada na maioria dos sistemas de informação das organizações. Nesse contexto, a aferição da qualidade de dados ganha ênfase nas organizações, pois percebe-se que a falta de qualidade é responsável por perdas de tempo, dinheiro e oportunidades, sendo necessária a criação de mecanismos para atenuar ou suprimir esse problema.

Torelli (2005) discute como os problemas de qualidade podem surgir, a Figura 1 ilustra uma representação de como ocorrem problemas de qualidade de dados em sistemas de informação. Esse autor discute que problemas relacionados à credibilidade podem surgir, por exemplo, quando ocorre falta de concordância entre as várias fontes de um mesmo dado, assim, os usuários da informação não reconhecem a fonte para a qual o problema pode ser atribuído, apenas percebe o conflito entre os dados apresentados. Além disso, quando ocorre o acúmulo desses conflitos, a integridade dos dados pode ser afetada, o que acarreta numa baixa reputação dos dados e, conseqüentemente, pouco valor é agregado ao sistema.

Torelli (2005) também alerta para a subjetividade no processo de produção dos dados, i.e. quando um dado é interpretado, sua qualidade em relação ao dado bruto é reduzida, comprometendo a sua objetividade. De acordo com o autor, "inicialmente, apenas aquele com conhecimento do processo de produção dos dados tem esses potenciais problemas". Ademais, o acúmulo de subjetividade na produção dos dados também resulta em informação de

credibilidade questionável e com baixa reputação e pouco valor agregado, reduzindo a sua utilização pelo cliente. Torelli (2005) observou que diferentes formas de resolução dos problemas podem ser adotadas. Uma das soluções é baseada nos sistemas computacionais, ignorando os processos de produção dos dados, porém, essa alternativa produz dados de baixa qualidade e integridade. As outras soluções são baseadas tanto nos processos de produção dos dados quanto nos sistemas computacionais, o que resulta numa melhora na qualidade dos dados. Essa observação respalda a necessidade de se tratar a qualidade dos dados na sua origem, isto orientou o desenvolvimento desta dissertação.

**Figura 1 - Representação da ocorrência de problemas de qualidade**



Fonte: Adaptado de Torelli (2005).

Olson (2003) destaca alguns fatores que causam problemas na qualidade dos dados:

- **Problemas ocasionados na entrada inicial de dados:** erros de digitação; seleção incorreta de itens em um formulário; valores corretos em campos errados; omissão do dado correto pelo usuário por considerar invasão de privacidade ou entrada de dados incorretos com objetivo de favorecer o usuário;



- **Degradação de dados:** ocorre com dados que estão corretos inicialmente, porém, com o seu processamento, tornam-se inadequados ou desatualizados;
- **Mudanças e reestruturação dos dados:** quando o processo de mudança e reestruturação de dados transforma dados corretos em dados imprecisos. Esse problema costuma ocorrer por falta de conhecimento total das bases de origem e/ou destino dos dados, ou quanto os sistemas não comportam a ocorrência de mudanças;
- **Uso de dados:** quando na inserção de dados em objetos de negócios o usuário não compreende o significado do dado ou o contexto em que esse se insere, essa interpretação equivocada pode gerar imprecisão dos dados.

## 1.2 JUSTIFICATIVA

Os repositórios de dados têm recebido maior atenção nos últimos tempos, com a criação de diferentes tecnologias de gerenciamento de bases de dados robustos, serviços de replicação de dados, além de suporte à segurança e *backup* (SALVADOR *et al.*, 2012). Wang *et al* (1995; 2000) propõem que a questão da qualidade da informação seja vista como uma correta adequação de sua utilização para o consumidor. Para tanto, ele propõe um conjunto de técnicas para a aplicação de um controle de qualidade total da informação, que ele denomina TDQM – *Total Data Quality Management*, definindo e aplicando algumas dimensões para análise ontológica dos dados (WANG *et al*, 1995)

Pinho (2001) apresenta uma ferramenta, AQUA, que, em conformidade com a ISO 9016 auxilia a avaliação da qualidade de dados pela não conformidade, mas não demonstra uma utilização automática para a aplicação das métricas propostas por Wang *et al* (2000) e Wang *et al* (1995). Em contrapartida, Amaral (2003) aplica em uma ferramenta, a AQUAWARE, os conceitos definidos por Wang *et al.* (2000) e Wang *et al.* (1995), em um ambiente de *Data Warehouse*. Contudo, vale salientar que de acordo com Salvador *et al.* (2012), apesar de muitas tecnologias de acesso aos dados terem sido criadas nestes anos, a exemplo de *Data Warehouse*, *Data Mart*, *Data Mining*, pouca coisa tem sido feita com relação ao dado propriamente dito.

A qualidade de dados vem, atualmente, ganhando ênfase nas organizações, a partir do momento em que se observa que dados com baixa qualidade podem ser responsáveis por custos que poderiam ser evitados, perdas de tempo e de oportunidades. Contudo,

Salvador *et al.* (2012) observa que em muitas organizações, a qualidade de dados ainda é uma questão que recebe baixa prioridade e deve ser entendida como um processo que requer constante avaliação e melhoria. Ademais, é observado que existem estudos acerca das questões da qualidade, baseados nos critérios propostos por Wang *et al* (2000), mas não é encontrada na literatura, que será discutida no Capítulo 4, uma aplicação prática e automatizada ou semiautomatizada desses conceitos em um ambiente transacional.

### 1.3 MOTIVAÇÃO

As corporações que utilizam sistemas ERP (*Enterprise Resource Planning*) para gerar informações e desejam medir a qualidade da informação gerada por seus sistemas, necessitam que a aplicabilidade das técnicas de TDQM sejam facilmente implementadas nas suas bases de dados. A inserção dessas técnicas em um sistema transacional, de maneira automatizada ou semiautomatizada, poderia reduzir os custos de implantação de um ambiente gerenciado para prover a qualidade desejada dos dados e a sua adoção por um número maior de organizações.

Portanto, a partir dessa suposição, duas hipóteses podem ser levantadas:

- a) É possível a criação de uma ferramenta que mensure a qualidade de dados em sistemas transacionais?

Através de uma pesquisa bibliográfica, acerca das definições de qualidade de dados e as técnicas para a certificação da qualidade propostas na TDQM, é possível se chegar a um modelo de processo para a criação e utilização de uma ferramenta que ajude a determinar o grau de qualidade do conteúdo contido em um banco de dados. Para isso, é necessária a verificação de propostas já realizadas na literatura e uma adequação às metodologias utilizadas no tratamento e certificação da qualidade em bases de dados.

- b) O uso dessa ferramenta melhoraria a qualidade do sistema de informação?

A qualidade de um sistema depende de diversos fatores que podem ser determinantes para a elevação ou diminuição do seu grau. Dentre esses fatores, um pode estar diretamente ligado à qualidade dos dados que ele gera/gerencia. Partindo desse pressuposto, é provável

que um sistema que gere informações confiáveis tenha um grau de qualidade maior que aquele que apresente informações não confiáveis como resultado de um processo.

Para responder essas questões e suprir a ausência de uma ferramenta que automatize as técnicas da TDQM para a análise de dados transacionais é proposto neste trabalho uma aplicação dos conceitos da TDQM a partir do desenvolvimento de uma ferramenta visual denominada FEMEQ – Ferramenta de Medição de Qualidade de Dados em Ambientes Transacionais. A avaliação do grau de qualidade dos dados contidos em uma base pode, então, ser executada, utilizando-se a ferramenta proposta por esse trabalho em um processo de avaliação da qualidade de dados para sistemas de informação. O processo de avaliação proposto é baseado na visão do usuário, por que este é o principal favorecido com os resultados oriundos do processamento dos dados em um sistema de informação e não se pode ter um serviço de alta qualidade quando o maior interessado neste serviço não estiver satisfeito (PRESSMAN, 2006). Wand & Wang (1996) afirmam que os usuários realizam ações baseadas em informações geradas pelos sistemas e que “o conceito de qualidade de dados ou informação depende do uso real dos dados. O que pode ser considerado um bom dado em um caso, (para uma aplicação específica ou usuário) pode não ser suficiente em outro caso”.

Para determinar se a qualidade da base de dados influenciará na qualidade do sistema como um todo, será realizada a aplicação da ferramenta proposta, baseada na concretização de uma pesquisa para determinar o seu impacto da qualificação dos dados frente às necessidades dos usuários, de acordo com os requisitos de cada um.

### 1.3 OBJETIVOS

#### **1.3.1 Objetivo Geral**

Este trabalho tem como objetivo apresentar um modelo para a avaliação da qualidade de dados em ambientes transacionais, baseado nas etapas propostas na TDQM (WANG et al, 2000), definido por um ambiente baseado em um processo de avaliação guiado pela visão da qualidade do usuário e apoiada por uma ferramenta que auxilie na tarefa de avaliação da qualidade dos dados em ambientes transacionais.

### 1.3.2 Objetivos Específicos

- a) Realizar uma revisão bibliográfica acerca da Qualidade de Dados e Informações e identificar na literatura ferramentas e propostas correlatas de forma que possa subsidiar a solução para o problema que se deseja resolver;
- b) Definir um modelo de processo com a identificação das dimensões e métricas para a análise da qualidade dos dados sob a perspectiva do usuário e suportado pela ferramenta proposta;
- c) Especificar e implementar uma ferramenta que permita a avaliação da qualidade de dados com base nas técnicas TDQM e na visão do usuário do sistema de informação;
- d) Realizar um estudo de caso para avaliação e validação do processo e ferramenta propostos.

### 1.4 Organização da dissertação

Além deste capítulo introdutório, esta dissertação está organizada em mais 6 (seis) capítulos, conforme apresentados a seguir:

- a) Capítulo 2 - Metodologia: no qual é apresentada a metodologia proposta para a elaboração desta dissertação;
- b) Capítulo 3 – Qualidade de Dados: capítulo em que são apresentados conceitos e definições fundamentais para a compreensão deste trabalho, como os conceitos de dado, informação e conhecimento, demonstrando suas diferenças e semelhanças;
- c) Capítulo 4 – Estado da Arte: são discutidos trabalhos que aplicaram métodos para realizar a avaliação da qualidade de dados, entendendo os critérios utilizados, pontos positivos e possíveis deficiências da metodologia utilizada nesses trabalhos;
- d) Capítulo 5 – Apresenta-se um ambiente para a medição da qualidade de dados em ambientes transacionais: são apresentadas uma ferramenta, FEMEQ e o modelo de processo para a criação e utilização da ferramenta

de apoio à medição de qualidade de software utilizando os conceitos da TDQM;

- e) Capítulo 6 – Estudo de caso: é apresentado um estudo de caso para avaliar a qualidade dos dados em diferentes sistemas de informação com o auxílio do processo e ferramenta propostos nesta dissertação como forma de validação das principais ideias propostas;
- f) Capítulo 7 – Conclusões e Trabalhos Futuros : são discutidas as conclusões e perspectivas de trabalhos futuros.

## CAPÍTULO 2 - METODOLOGIA

No presente capítulo será apresentada a metodologia proposta para a elaboração deste trabalho. A metodologia utilizada baseia-se no método estruturalista (MARCONI; LAKATOS, 2003) e foi dividido em quatro etapas: a Bibliográfica; Experimental e Pesquisa com *Survey*. A classificação da pesquisa é apresentada no Quadro 1, e será discutida nas seções seguintes.

**Tabela 1 - Quadro Metodológico**

Metodologia da Pesquisa	Bibliográfica Pesquisa com <i>Survey</i> Experimental
Natureza	Aplicada
Abordagem	Qualitativa
Objetivos	Exploratória Pesquisa Descritiva
Problemas Identificados	Baixa qualidade de sistemas decorrentes da baixa qualidade de dados. Carência de ferramenta de medição de qualidade de dados em ambientes transacionais.
Variáveis	Independentes: Metodologia de medição de qualidade e modelo de métricas. Dependentes: Melhoria da qualidade de sistemas e mensuração da qualidade.

Fonte: o autor

### 2.1 MÉTODO ESTRUTURALISTA

A pesquisa deste trabalho baseia-se no Método Estruturalista, partindo do pressuposto que um sistema pode apresentar má qualidade, dentro de parâmetros pré-definidos por um usuário desse sistema. Esse método permitirá uma visão abstrata das questões que norteiam

problema de qualidade, através da investigação de modelos de qualidade e permitindo aplicação concreta desses modelos na resolução das questões analisadas.

## 2.2 METODOLOGIA DA PESQUISA

### 2.2.1 Pesquisa Bibliográfica

Deve ser realizada a identificação de referenciais teóricos acerca de qualidade de dados e sua aplicação através de artigos científicos e livros publicados, a fim de conhecer o que já foi pesquisado e o que está sendo realizado no que tange à questão da qualidade de software.

A revisão bibliográfica foi realizada a partir das duas questões de investigação, discutidas no Capítulo 1, cujas respostas nortearam a proposta e desenvolvimento da ferramenta de medição de qualidade de dados em sistemas transacionais, como se segue:

- a) A qualidade de dados é um fator impactante para a qualidade de um sistema transacional?
- b) É possível desenvolver uma ferramenta que realize a medição dos dados em ambientes transacionais?

Com base nessas questões foram definidas as estratégias de busca e as fontes de pesquisa:

#### a) Estratégias de Busca

Para a realização do material teórico acerca do tema trabalhado, os termos para uma consulta nos motores de busca da Web foram criados com os seguintes critérios:

- Foram identificados os termos a partir das questões definidas no processo de investigação;
- Foram escolhidos trabalhos que tratem diretamente dos temas: qualidade de dados, dimensões de qualidade de dados e ferramentas de avaliação da qualidade de dados;
- Os termos foram traduzidos para o inglês, a fim de ampliar os resultados nas pesquisas;

- Foram realizadas consultas em motores de buscas selecionados com os termos identificados.

**Tabela 2 - Termos de busca na Web.**

<b>Termos em Português</b>	<b>Termos em Inglês</b>
“Qualidade de Dados”	“Data Quality”
“Qualidade da Informação”	“Information Quality”
“Dimensões da Qualidade de Dados”	“Data Quality Dimensions”
“Fundamentos de Qualidade de Dados”	“Foundations of the Data Quality”

Fonte: o autor.

### **b) Fontes de Pesquisa**

Para o referencial teórico foram realizadas consultas, utilizando os seguintes mecanismos de busca: Google (<http://www.google.com>); IEEEExplore (<http://ieeexplore.ieee.org/Xplore/home.jsp>) e Mendeley (<http://www.mendeley.com/research-papers/>)

Para obter o referencial com qualidade acerca do tema dissertado, foram utilizadas as palavras de busca definidas no Quadro 2, com os termos em inglês, o que resultou em diversos resultados, que foram sendo refinados seguindo os seguintes critérios:

1. Os trabalhos foram priorizados, por relevância, na seguinte ordem de classificação: material (artigos, dissertações e teses) publicados em congressos e em revistas científicas;
2. Foi realizada leitura dos resumos dos trabalhos, a fim de verificar se o tema abordado apresentava proposta semelhante ao presente estudo, sendo, ainda, analisadas as fontes de pesquisas, referencial bibliográfico, dos trabalhos selecionados.

Foram excluídos da pesquisa: (i) todos os materiais que não tratem primária ou secundariamente acerca de: qualidade de dados, dimensões de qualidade de dados e ferramentas de avaliação da qualidade de dados, (ii) os materiais repetidos, (iii) materiais que não tenham, claramente, relevância para a pesquisa, i.e., aqueles que tratam do tema proposto mas não agregam nada novo à pesquisa, (iv) além daqueles que não são de livre acesso para consulta na Web.

Nessa fase da pesquisa foi gerado o referencial bibliográfico deste trabalho.



### **2.2.2 Pesquisa com Survey**

No Capítulo 6, a aplicação de um questionário foi submetida aos gestores das empresas que participaram no estudo de caso apresentado por este trabalho e tiveram a qualidade dos dados dos seus sistemas de informação avaliados, a fim de obter informações acerca da definição da qualidade de dados sob a sua ótica, auxiliando na determinação das dimensões de qualidade para a pesquisa experimental a ser realizada com o auxílio da ferramenta FEMEQ, proposta neste trabalho, e permitindo a identificação e definição das métricas para a qualidade dos dados gerados pelo sistema transacional da empresa.

### **2.2.3 Pesquisa Experimental**

Depois de realizada a pesquisa bibliográfica, a ferramenta para a medição da qualidade dos dados foi implementada baseada nos conceitos de qualidade de dados definidos por Wang *et al* (2000), como pode ser visto no Capítulo 5 e, a partir daí, começou-se a fase de pesquisa experimental, na qual a ferramenta criada passou a ser testada em um sistema real, apresentado no Capítulo 6, neste caso, um sistema de gerenciamento de hotéis, um de academias e outro de clubes, medindo, a cada transação, a qualidade dos dados gerados, de acordo com o definido pelos gestores das empresas, a fim de determinar se a qualidade desses dados influenciaram na qualidade do sistema.

Durante essa fase da pesquisa foram gerados os seguintes artefatos:

- A ferramenta de mensuração da qualidade de dados;
- Relatório atestando a qualidade ou não dos dados gerados pelo sistema.

### **2.2.4 Natureza Aplicada**

Basicamente, em sua natureza, este trabalho se baseou em referenciais teóricos já publicados, aplicando metodologias e técnicas pré-definidas, resultando em uma aplicação prática de tais técnicas em uma ferramenta visual que auxilie a avaliação da qualidade dos dados em sistemas transacionais.

### **2.2.5 Abordagem Qualitativa**

Com caráter de pesquisa para entendimento de um fato, este trabalho realiza uma abordagem qualitativa, na qual o objeto de estudo é a relação entre a qualidade de dados e a qualidade dos sistemas. Portanto, o universo de amostra para testes é relativamente pequeno, no qual a ferramenta gerada foi testada em algumas empresas, portadoras de sistemas de informações.

Para tanto, todos os sistemas testados possuem as mesmas características, para que a avaliação dos sistemas retorne dados confiáveis e homogêneos, como se segue:

- I. Todos os software operam como sistemas transacionais
- II. Possuem arquitetura Cliente-Servidor
- III. O ambiente de execução é desktop para sistemas operacionais Windows
- IV. Todos os sistemas utilizam o sistema gerenciador de banco de dados relacional Firebird 2.1

### **2.3 Considerações finais**

No presente capítulo foi apresentada a metodologia utilizada para o desenvolvimento desta dissertação, demonstrando a sua natureza, abordagem, objetivos, problemas identificados e variáveis. A aplicação dessa metodologia possibilitou a realização de uma pesquisa estruturada e organizada. No capítulo seguinte serão discutidos alguns conceitos importantes para o desenvolvimento deste trabalho.

## CAPÍTULO 3 - QUALIDADE DE DADOS

Neste capítulo são apresentados conceitos e definições fundamentais para a compreensão deste trabalho. São definidos os conceitos de dado, informação e conhecimento, demonstrando suas diferenças e semelhanças, para se discutir o que se entende por qualidade de dados, as dimensões de qualidade e qualidade de software.

### 3.1 DADO, INFORMAÇÃO E CONHECIMENTO

Comumente, as pessoas consideram o próprio dado como sendo uma informação. No entanto, Zins (2007) afirma que tanto a comunidade científica quanto as organizações têm conceituado ambos os termos de maneiras diferentes, baseando-se nas visões de seus campos de estudo e do contexto. Para o referido autor, os dados podem ser definidos como um conjunto de sinais que representam estímulos aos sentidos ou percepções empíricas. Para Kossowski (2010), o dado é um registro do evento, que, para se aproximar da realidade, demanda um maior nível de detalhamento possível, enquanto Bell (1999) considera que um dado é uma sequência ordenada de eventos ou estatísticas. Sendo assim, os dados podem ser considerados como uma sequência de elementos quantificados ou qualificáveis que, quando analisado isoladamente, não consegue transmitir uma mensagem ou representar algum conhecimento. Para este trabalho, o dado será considerado como o valor de um atributo observável, mensurável ou calculável que pode ser descrito através de representações formais, através de uma estrutura pré-definida, livres de qualquer definição semântica, podendo ser processada e armazenada em computador. Os dados são elementos que constituem a matéria prima da informação.

Já a informação deriva dos dados. É o resultado da organização lógica ou transformação que os dados sofrem, através de categorização, correção, cálculo ou contextualização, a fim de gerar deduções lógicas confiáveis. Em oposição aos dados, ela exige análise e deve possuir um significado em relação ao seu contexto, permitindo ao seu utilizador, gerar conhecimento acerca do que está sendo tratado. Quando transformado em informação, o dado torna-se insumo básico das empresas, estando presente em todas as etapas do processo, desde o início até o *Feedback*, que permite o carregamento de novos dados no sistema (KOSSOWSKI, 2010). Pipino *et al.* (2002) afirmam que dados e informações são

também comumente interpretados como sinônimos, vez que é o processamento dos dados que gera informações.

Contudo, segundo Malhorta (1998), existe uma confusão entre conhecimento e informação. Para o autor, a informação, gerada por computadores não permite a interpretação humana, enquanto o conhecimento está em um contexto mais subjetivo de ação do usuário, além de basear-se na informação. Ademais, o processo de transformar os dados em informação depende do conhecimento, que deriva de dedução e assimilação de informações em um contexto percebido, experiências ou regras organizacionais (DE PAULA, 2009). Dessa forma, o conhecimento cresce graças a um constante processo de organização e filtragem das informações. Oliveira & Amaral (1999) consideram que o conhecimento ocorre quando alguém interpreta a informação, utiliza-a em seu processo de decisão e identifica qual informação será útil nesse processo.

### 3.2 QUALIDADE DE DADOS

Para compreender os fundamentos que embasam a ideia de qualidade de dados, é necessário definir, em primeira instância, o que se entende por qualidade. De acordo com Kossowski (2010), são diversos os conceitos sobre qualidade, expressão utilizada desde muito tempo, porém, tratando especificamente de qualidade em produtos. Evans & Lindsay (1999) consideram a qualidade um conjunto de traços e atributos de determinado produto ou serviço que busca alcançar ou exceder as expectativas e necessidades do cliente. Pinho (2001) considera que a obtenção da qualidade exige mudança de postura e paradigmas dentro de uma empresa, tendo sempre como objetivo a satisfação do cliente. Tal mudança de postura, ainda de acordo com a autora, envolve a percepção de tratar as respostas aos problemas como uma oportunidade de alcançar a confiança e satisfação do cliente, o que pode garantir, em última instância, um aumento na participação do mercado, bem como a fidelidade dos clientes. Percebe-se, portanto, que apesar de ter um conceito complexo, já que depende da perspectiva do avaliador (PINHO, 2001), é consenso que a qualidade busca a satisfação de todos os indivíduos envolvidos no processo, desde a empresa até o cliente, e está presente nas características intrínsecas dos objetos, bem como na aplicação desses.

Com a globalização e crescimento da tecnologia na área da informatização, cada vez mais as empresas possuem uma quantidade sempre crescente de dados sobre seus clientes, fornecedores, produtos e negócios, gerando, muitas vezes, grandes bancos de dados que serão

alicerces para a tomada de decisões nos negócios dessas organizações. A economia passou ser embasada no conhecimento, o que exige agilidade nas tomadas de decisões, na capacidade de transformar informações em conhecimento e estes em ações de negócio, para garantir vantagem competitiva. O que Cavalcanti-Gomes (2000) chama de Inteligência Empresarial. Drucker (1993) *apud* Cavalcanti-Gomes (2000), afirma que a “questão central para o executivo moderno é ser capaz de usar o conhecimento para criar novos produtos e serviços”. Tais conhecimentos, na atualidade, provêm de modernos sistemas de informação que conseguem, rapidamente, responder às questões administrativas das organizações. Para isso, elas fazem uso dos dados contidos em seus bancos sempre que precisam criar campanhas de marketing, onde é imprescindível o envio de malas diretas, entender o comportamento e necessidade de seus clientes, conhecer seus produtos, realizar pedidos a fornecedores, efetuar cobranças, entre outros diversos procedimentos administrativos necessários à vida dessas empresas. Kossowski (2010) assevera que um sistema de qualidade pode ser compreendido como um sistema com informação consistente, que permite a obtenção de respostas acuradas que otimizem o planejamento e a rentabilidade da organização por meio da satisfação do mercado consumidor. No entanto, tais bases podem estar cheias de lixo, dados sem sentido, incompletos ou mal organizados, o que acaba afetando o resultado final dessas operações. Contudo, o conceito de qualidade de dados é relativo, sendo melhor entendido no sentido de “adequação ao uso”, já que a qualidade de dados de um sistema pode ser suficiente em um contexto mas ser inadequada em outro (SALVADOR *et al.*, 2012).

Batini (1998) afirma que uma baixa qualidade de dados pode afetar seriamente ou, até mesmo, causar sérios danos à eficiência e efetividade dos negócios e organizações. Para o autor, dados de alta qualidade são aqueles que estão de acordo com os requisitos de uso e cumprem os papéis a que foram destinados. Salvador *et al.* (2012) apontam que a baixa qualidade dos dados gerados e os custos envolvidos no reprocessamento destes não são tão óbvios para muitas empresas, que acabam, ainda, perdendo quantidade significativa de dinheiro, tempo, e, conseqüentemente, oportunidades. Ainda de acordo com os autores, é comum que atividades de correção e reprocessamento de dados façam parte do cotidiano das empresas que ainda não atentaram para a magnitude do impacto desse problema nas organizações.

MacGilvray (2008) defende que a qualidade dos dados é o grau em que as informações geradas por esses dados podem ser consideradas confiáveis, não importando o quanto o usuário irá buscá-los. O cuidado com a base de dados dos clientes proporciona à empresa uma maior oportunidade de negócios, haja vista que haverá uma melhor gestão no contato com

esses clientes, a baixo custo, proporcionando um aumento de receita. Periodicamente, decisões são tomadas, baseadas em análise de informações resultante de consultas a bases de dados, sem mesmo que se saiba a qualidade do conteúdo ali presente, o que aumenta enormemente o risco de resultados inadequados, em função da não conformidade de conteúdo. Por esse motivo, a qualidade de dados vem ganhando notada atenção frente às organizações, haja vista que adotar uma estratégia para garanti-la vem se tornando fundamental no mundo dos negócios (MACGILVRAY, 2008)

Willshire e Meyen (1997 *apud* PINHO, 2001) apresentam dois tipos de avaliações realizadas em um banco de dados, a avaliação quantitativa, que é objetiva porque é comumente realizada por *software* que se baseiam em indicadores para medir a qualidade da representação e da estrutura de um banco de dados; e a avaliação qualitativa, que é subjetiva, já que são especialistas ou usuários do banco de dados que a realizam, ou também ferramentas automatizadas, mas que utilizam indicadores subjetivos para qualificar um banco de dados e sua capacidade de uso.

De acordo com Pinho (2001), a avaliação quantitativa considera como essenciais as dimensões de qualidade *precisão* e *objetividade*. Contudo, a avaliação objetiva por si só não é suficiente para classificar os dados como de "alta qualidade", além de grande parte dos usuários dos dados considerarem que tais dimensões também não são suficientes para a avaliação (AMARAL, 2003). Já na avaliação qualitativa, os indicadores subjetivos se configuram como características de qualidade pré-definidas que tentam perceber a expectativa e análise do usuário sobre a qualidade de dados. A autora assevera ainda que uma vez que as exigências e necessidades mudam ao longo do tempo, assim também o faz a ideia de qualidade, fazendo com que as soluções de avaliação passem por constantes mudanças, o que, ainda assim, é mais vantajoso para uma empresa do que corrigir possíveis erros oriundos de dados de baixa qualidade.

### 3.3 TOTAL DATA QUALITY MANAGEMENT - TDQM

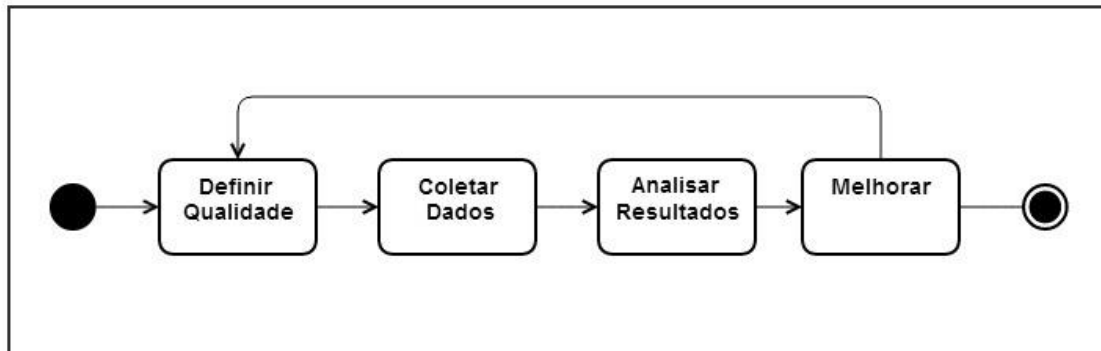
Implantar um programa de gerenciamento de qualidade de dados não é uma tarefa trivial. Uma vez que o conceito de qualidade de dados é relativo (TAYI, 1998), o que é bom em termos de qualidade para uma empresa, pode não ser suficiente para outra. Para muitos, quando se fala em qualidade de dados, a questão da acurácia (precisão) já é o suficiente, porém existem algumas outras dimensões que também devem ser observadas. Dentre elas, algumas não menos importantes são completeza, valor, consistência, acessibilidade e

temporalidade. As diversas dimensões da qualidade serão ainda tratadas neste trabalho. Dentro desse contexto, cada caso de aplicação de regras para a mensuração da qualidade de dados deve ser estudado individualmente, adequando regras e características de qualidade em acordo com a realidade e cultura de cada empresa.

O *Total Quality Management* – TQM representa mais do que uma abordagem ao problema de gestão da qualidade. Oliveira e Amaral (1999) o definem como uma "filosofia baseada no princípio da melhoria contínua, na crença de que é sempre possível fazer mais e melhor [...] uma abordagem que integra técnicas de gestão com ferramentas da qualidade." Wang *et al* (1995, 2000) trazem uma aplicação do conceito TQM para a qualidade dos dados, denominada *Total Data Quality Management* - TDQM, que se baseia no produto, no ciclo de melhoria contínua e na gestão do processo. Ademais, Wang *et al.* (2000) simplificam o entendimento ao relacionar o processo de manufatura de produtos e manufatura de informações, i.e. enquanto a manufatura de produtos transforma materiais brutos em produtos propriamente ditos, a manufatura da informação transforma os dados brutos em um produto de informação. Essa concepção dos dados como sendo um produto, facilita a avaliação de sua qualidade por meio do TDQM (SOUZA, 2009). Contudo, é importante observar que uma das principais diferenças entre a qualidade de um produto e a qualidade de informação, deve-se ao fato de que enquanto a qualidade do produto pode ser medida por indicadores físicos, como as especificações do projeto, a qualidade da informação possui parâmetros mais subjetivos, pois são embasados na perspectiva do usuário (AL-HAKIM, 2004).

Para o desenvolvimento do TDQM são estabelecidas quatro etapas: (i) elencar as dimensões de qualidade a serem adotadas; (ii) medir a qualidade ao longo do "ciclo de vida" da informação; (iii) verificar possíveis causas dos problemas de qualidade e, finalmente, (iv) implementar melhorias nos processos, a fim de manter a qualidade dos dados, prevenindo, também, contra vulnerabilidades futuras (WANG *et al.*, 2000), como mostra Figura 2.

**Figura 2 - Etapas para a Aplicação da TDQM**



Fonte: Adaptado de (Wang *et al.*, 2000)

De acordo com a Figura 2, cada fase depende da anterior para a sua execução, sempre voltando à primeira, num ciclo evolutivo, para se manter a qualidade. A descrição das etapas segue abaixo:

- a) Definir Qualidade: A definição da qualidade é a fase em que o avaliador utilizará seus conhecimentos acerca das questões de qualidade e das regras de negócio do SI a ser avaliado e definirá uma metodologia de aplicação dessas dimensões no sistema em avaliação;
- b) Coletar Dados: Nessa fase, os dados são avaliados de acordo com a metodologia definida na fase anterior e preparados para serem analisados;
- c) Analisar Resultados: Depois de coletados os dados, é necessário realizar uma análise dos resultados obtidos durante a fase de coleta, realizando uma avaliação para certificar a qualidade dos dados gerados;
- d) Melhorar: Nessa fase é realizado um estudo acerca dos processos que apresentem problemas de qualidade na geração das informações e é realizado um plano de intervenção para tentar resolvê-lo. Concepção.

Para caracterizar a qualidade dos dados, uma estrutura hierárquica pode ser utilizada. Entende-se que os dados possuem características que podem ser avaliadas e medidas, elas são denominadas atributos da qualidade. Tais atributos são geralmente agrupados em dimensões da qualidade, as quais relacionam os atributos que representam uma compreensão sobre a qualidade de dados. Esses atributos são avaliados em conjunto. Já as dimensões são, então, agrupadas em categorias, que facilitam a identificação de padrões na ocorrência de problemas relacionados com a qualidade. Ademais, o estabelecimento de métricas, em associação com as



dimensões de qualidade, permite a obtenção de medidas, que indicarão o grau de presença dos critérios nos dados (AMARAL, 2003).

### 3.4 QUALIDADE DE SOFTWARE

O conceito de qualidade de *software* surgiu da demanda por organização e padronização no desenvolvimento de *software*, que eram criados sem critérios de qualidade bem definidos e tinham baixa confiabilidade, além de uma manutenção complexa, quando realizada por outra pessoa que não participou do seu desenvolvimento (PRESSMAN, 2006). A qualidade de *software* pode ser definida como sendo um conjunto de propriedades a serem satisfeitas, em determinado grau, de modo que o *software* satisfaça às necessidades de seus usuários (ROCHA, 1987 *apud* PINHO, 2001).

A melhoria da qualidade de *software* é dividida em melhoria de qualidade de processo de *software*, que avalia os requisitos necessários ao consumidor, instala e realiza a manutenção do produto, e a melhoria de qualidade de produto de *software*, que avalia o produto já finalizado e define quais características um produto com qualidade deve ter, bem como os meios de avaliar essas características. Exemplos de normas para a melhoria da qualidade do processo são a ISO/IEC 9000, a ISO/IEC 12207, o modelo de maturidade CMMI, o SPICE, além do MPS.BR, como demonstrado por PRESSMAN (2006) . Já as normas que avaliam a qualidade do produto de software são a ISO/IEC 14598, ISO/IEC 12207 e a ISO/IEC 9126.

A norma ISO/IEC 9126, publicada em 1991, busca a qualidade do produto de *software* e apresenta um conjunto de atributos descritores de qualidade de qualquer produto de software: (1) funcionalidade; (2) confiabilidade; (3) usabilidade; (4) eficiência; (5) manutenibilidade e; (6) portabilidade. Uma modificação feita na norma incluiu a subcaracterística "conformidade" em todas as características principais, porque em todas elas é possível notar a busca por seguir a legislação, aos padrões internos e demais normas associadas a essas características (ISO/IEC 9126-1, 2001).

Os atributos de qualidade de software podem ainda ser classificados como funcionais ou não funcionais. Nesse caso, os atributos funcionais são aqueles relacionados com "o que" deve ser feito; se aplicam a partes do software e módulos do sistema, enquanto os atributos não funcionais podem se aplicar a qualquer produto, em qualquer fase do processo de

desenvolvimento, como especificações, códigos ou manuais, além de estar relacionados com o "quão bem" deve ser feito (McDERMID, 1994 *apud* BUENO, 2006).

### 3.5 DIMENSÕES DA QUALIDADE

Wand & Wang (1996) afirmam que os defeitos de qualidade dos dados obtidos são usados para definir as dimensões de qualidade. A Tabela 3 apresenta as definições das 16 dimensões de qualidade, de acordo com Pipino *et al.* (2002).

Wand & Wang (1996) afirmam que a escolha das dimensões de qualidade é inicialmente intuitiva, por ser baseada na experiência industrial ou revisão de literatura. Porém, os autores destacam ainda que não parece haver um consenso sobre as dimensões de qualidade de dados.

Wang & Strong (1996), visando identificar aspectos de qualidade relevantes sob o ponto de vista dos consumidores dos dados, apresentam uma divisão das dimensões de qualidade em quatro categorias: (1) Intrínseca, com as dimensões acurácia, objetividade, confiabilidade e reputação; (2) Acessibilidade, com as dimensões acessibilidade e segurança; (3) Contextual, com as dimensões relevância, valor agregado, temporalidade, completeza e quantidade apropriada de dados e; (4) Representacional, com as dimensões facilidade de interpretação, facilidade de entendimento, consistência na representação e concisão. Favaretto (2005) explica as características de cada categoria. De acordo com o autor, a categoria "Intrínseca" demonstra a qualidade inerente de cada informação, as categorias "Acessibilidade" e "Representação" expressam a importância do papel dos Sistemas de Informação e a categoria "Contextual" mostra a inserção da informação no meio em que é utilizada.

De acordo com Wang *et al.* (1993), para a gestão da qualidade de dados é necessário entender quais as dimensões da qualidade são essenciais para o usuário. A credibilidade, por exemplo, pode refletir a avaliação de um indivíduo da confiabilidade da fonte de dados, a comparação com um padrão comumente aceito, e experiência anterior. Já a dimensão "quantidade apropriada de dados" deve garantir que a quantidade de dados esteja na medida certa, sendo nem tão pouco, nem demais. Uma métrica geral que incorpora esta compensação baseia-se em duas proporções simples: a razão entre o número de unidades de dados fornecidos para o número de unidades de dados necessários, e a razão entre o número de unidades de dados necessários para o número de unidades de dados fornecidos (PIPINO *et al.*, 2002).

Tabela 3 - Dimensões da qualidade

<b>Dimensão</b>	<b>Definição</b>
<b>Acessibilidade</b>	a medida em que os dados estão disponíveis ou facilmente e rapidamente recuperáveis
<b>Quantidade apropriada de dados</b>	a medida em que o volume de dados é apropriado para a tarefa em mãos
<b>Credibilidade</b>	a medida em que o dado é considerado como verdadeiro e confiável
<b>Completeza</b>	a medida em que o dado não está faltando e tem amplitude e profundidade suficiente para a tarefa
<b>Representação concisa</b>	a medida em que os dados são representados de forma compacta
<b>Representação consistente</b>	a medida em que os dados são apresentados no mesmo formato
<b>Fácil de manipular</b>	a medida em que os dados são fáceis de manipular e aplicar a diferentes tarefas
<b>Livre de erro</b>	a medida em que os dados são corretos e confiáveis
<b>Interpretabilidade</b>	a medida em que os dados estão em linguagens, símbolos e unidades apropriadas e as definições são claras
<b>Objetividade</b>	a medida em que os dados não são enviesados, são livres de julgamento precoce e imparciais
<b>Relevância</b>	a medida em que os dados são aplicáveis e úteis para a tarefa em mãos
<b>Reputação</b>	a medida em que os dados são altamente considerados em termos de sua fonte ou conteúdo
<b>Segurança</b>	a medida em que os dados são restritos de forma adequada para manter a sua segurança
<b>Temporalidade/opor- tunidade</b>	a medida em que os dados são suficientemente atualizados para a tarefa proposta
<b>Facilidade de entendimento</b>	a medida em que os dados podem ser facilmente compreendido
<b>Valor adicionado</b>	a medida em que os dados são benéficos e proporcionam vantagens pela sua utilização

Fonte: Pipino *et al.* (2002).

Dificuldades no tratamento dos dados podem surgir quando eles são normalmente gerados de diferentes maneiras e por diferentes razões, assim, à medida que centenas ou milhares de registros oriundos de fontes cada vez mais díspares são inseridos, o conhecimento das dimensões de qualidade de dados pode ser desconhecido ou perdido. Assim, o conhecimento acerca das dimensões auxilia o controle da qualidade de dados e o *design* de sistemas (WANG *et al.*, 1993). Pipino *et al.* (2002) afirmam que uma dificuldade frequente é definir as dimensões com precisão ou os aspectos de cada dimensão que se relacionam com a aplicação específica que se pretende utilizar. Ademais, Lima *et al.* (2009) observaram que alguns dos estudos analisados apresentaram discordância no entendimento do significado de algumas dimensões, chamando atenção para a necessidade de uma padronização de conceitos antes de iniciar qualquer processo de análise de dados. Dessa forma, no presente estudo, serão utilizados os conceitos apresentados por Pipino *et al.* (2002).

### 3.6 CONSIDERAÇÕES FINAIS

No presente capítulo foram discutidos conceitos importantes e necessários para o desenvolvimento desta dissertação. Observou-se que embora muitos autores considerem dado e informação como sendo sinônimos, outros consideram que os dados constituem a matéria prima da informação, que, por conseguinte, é resultado da categorização, correção e contextualização dos dados. Já o conhecimento, vai ser formado quando ocorre interpretação da informação e consequente utilização. Ademais, buscou-se definir qualidade de dados, conceito bastante relativo e complexo, uma vez que dependerá da circunstância e aplicabilidade dos dados, bem como do usuário que o utiliza. Contudo, é consenso que a qualidade busca a satisfação de todos os indivíduos envolvidos no processo, e que a baixa qualidade de dados tem gerado prejuízos financeiros e dificuldade na tomada de decisões em grandes organizações.

O conceito de TDQM, proposto por Wang *et al.* (1995, 2000), visa a gestão da qualidade dos dados, concebendo-os como um produto e relacionando o processo de gestão à manufatura de produtos, compreendendo, porém, que a qualidade da informação possui parâmetros mais subjetivos, já que tem como base a perspectiva do usuário. Também é discutido o conceito de qualidade de software, bem como as normas ISO que trazem parâmetros para avaliar a qualidade de produto de software. Por fim, os conceitos das

dimensões de qualidade de dados, propostos por Pipino *et al.* (2002), nortearão o presente trabalho.

## CAPÍTULO 4 - ESTADO DA ARTE

Neste capítulo pretende-se compreender o desenvolvimento de metodologias para avaliação da qualidade dos dados. Dessa forma, serão apresentados e discutidos trabalhos que aplicam métodos para realizar a avaliação da qualidade de dados, entendendo os critérios utilizados, características positivas e possíveis deficiências da metodologia aplicada. Ademais, serão apresentados estudos que avaliaram a qualidade de dados ou informação em diferentes áreas, demonstrando a necessidade de manter dados de qualidade em diferentes domínios.

### 4.1 METODOLOGIAS DESENVOLVIDAS PARA AVALIAÇÃO DA QUALIDADE DOS DADOS

Kovac *et al* (1997) descreveram a experiência de uma empresa de informação no desenvolvimento de um programa de TDQM para aprimorar a qualidade da informação, com o objetivo de incentivar e denotar a importância da gestão da qualidade em qualquer instância empresarial. O processo foi constituído por algumas etapas, primeiramente pela identificação do estado atual e do estado final desejado, bem como os benefícios e o escopo do projeto, além de um cronograma de atividades. Posteriormente, a equipe buscou verificar as necessidades do cliente, focando nas relações cliente/fornecedor interno.

Nesse trabalho foram consideradas como dimensões fundamentais para análise a precisão e a pontualidade de entrega. Assim, um conjunto de métricas foi desenvolvido para avaliar essas dimensões, por meio de um processo interativo denominado RUMBA, que identificava as relações cliente/fornecedor e definia as métricas. O processo foi baseado em cinco critérios: se a métrica era (1) razoável, (2) compreensível, (3) mensurável, (4) confiável e (5) alcançável. Assim, para cada relação cliente/fornecedor métricas eram identificadas.

Para a coleta e listagem das métricas de qualidade de dados foram utilizados três níveis de indicadores, os de processo, de qualidade e de medidas totais do sistema. Dessa forma, com base no processo RUMBA, cada equipe de cliente/fornecedor foi definida no processo de entrega, e as métricas relativas à pontualidade e precisão de entrega foram recolhidas e reportadas.

Em adição ao sistema de medição interna, um sistema de medição foi desenvolvido para quantificar o desempenho do sistema de fornecimento, a partir do ponto de vista do

cliente. De acordo com Kovac *et al.* (1997), o sistema foi integrado ao processo de produção, sendo referido como TRAQ (*Timeliness + Reliability + Accuracy = Quality*). Os autores acrescentam que os dois principais objetivos do sistema TRAQ são: (i) fornecer informações objetivas, medição consistente da qualidade dos dados e confiabilidade da entrega; (ii) proporcionar a melhoria contínua do processo de entrega, o que exige a especificação da razão e do local onde ocorreu o problema. Tal informação sobre as falhas permite aos especialistas desenvolverem soluções que irão prevenir a ocorrência de problemas futuros.

O processo TRAQ é constituído de seis fases: (1) processo de entrega de documentos, (2) problemas de rastreio, (3) relação entre processo e problema, (4) identificar a causa raiz, (5) desenvolver e implementar soluções e (6) elaboração de relatório de desempenho. Dados de tendências de erros do sistema TRAQ são usados para informar aos clientes sobre a ocorrência de erros, o que ajuda a criar bases para melhoria do relacionamento fornecedor/cliente. Além disso, os relatórios TRAQ incluem informações sobre quais áreas do processo são mais propensas a erro, assim, os especialistas funcionais analisam e identificam o maior número de processos passíveis de erro, priorizando os recursos internos de desenvolvimento para resolução dos problemas. Apesar de não apresentarem um software como ferramenta para a medição em tempo real para a medição da qualidade, Kovac *et al.* (1997) demonstraram que a implementação de um programa de gestão da qualidade total dos dados proporcionou notório benefício na integração de métricas em todos os níveis e processos dentro da cadeia de fornecimento. Para os autores, sem as métricas a empresa não teria identificado os problemas nos processos, priorizado soluções, garantido recursos corporativos para projetos de reengenharia ou determinado a eficácia dos esforços para o estabelecimento da qualidade total.

Lee *et al.* (2001) desenvolveram uma metodologia para avaliação da qualidade de informação, denominada AIMQ, para formar uma base para a avaliação de Qualidade de Informação (QI) e aferição. Os autores ilustraram a utilização da metodologia aplicando-a em cinco grandes organizações. A metodologia AIMQ é formada pelos seguintes componentes: (1) Modelo PSP/IQ (LEE et AL,200); (2) Instrumento IQA (*Information Quality Analysis*) (LEE et AL,200); (3) Técnicas de análise de lacunas de QI. O PSP/IQ é um modelo que organiza as principais dimensões de QI para que as decisões significativas sejam feitas com foco na melhoria da qualidade da informação, sendo essas dimensões desenvolvidas a partir da perspectiva dos consumidores de informação, e não de uma escolha lógica. O modelo PSP/IQ, baseia-se em quatro quadrantes: ruído, informação confiável, informação útil e informação utilizável. Os quadrantes representam aspectos da QI que são relevantes para as

decisões de melhoria da QI. O Instrumento IQA mede a qualidade da informação para cada uma das dimensões, sendo que essas medidas são mensuradas para formar medidas para os quatro quadrantes. As técnicas de análise de lacunas de QI avaliam a qualidade da informação de uma organização para cada um dos quatro quadrantes, e formam a base de concentração dos esforços para as melhorias. O modelo PSP/IQ vai de acordo com os princípios do TQM, focando no produto ou serviço e em como a qualidade pode ser avaliada pelas especificações ou expectativas do cliente e utiliza aspectos de qualidade que são relevantes para obter informações de melhor qualidade.

Para avaliar a qualidade da informação no nível de dimensão, as dimensões foram agregadas nos quadrantes PSP/QI. Valores para cada quadrante foram computados como a média dos valores de suas dimensões constituintes.

Para a análise foram utilizados os valores dos quadrantes como entrada e um conjunto de algoritmos para analisar e comparar as avaliações de QI e do instrumento QIA, bem como do modelo PSP/IQ. Esses valores são então utilizados para aferir a qualidade da QI de uma organização e focar no quadrante que mais necessita de melhoria. Para detectar áreas problemáticas de QI também se utilizou a técnica da análise de lacunas, que começa analisando um quadrante específico do modelo PSP/IQ, como a usabilidade, continuando para os outros quadrantes. Após análise individual de todos os quadrantes, os quatro são comparados para detectar padrões comuns e identificar áreas que necessitam de melhorias.

Lee *et al* (2001) consideram que a metodologia AIMQ é uma ferramenta prática de análise de QI. A metodologia foi útil na identificação de problemas de QI, chamando atenção para as áreas mais carentes de melhoria de qualidade, além de possibilitar o monitoramento da qualidade ao longo do tempo. Nesse trabalho, também não é apresentado nenhum software como ferramenta para a medição da qualidade dos dados utilizando a metodologia apresentada.

#### **4.2.1 Avaliação da qualidade pela não conformidade**

Pinho (2001) apresentou um modelo qualitativo, que se subdivide em três etapas, para avaliação da qualidade de dados baseada em não conformidades, compatível com a norma ISO/IEC 9126. A primeira etapa buscou estabelecer a importância de cada característica da qualidade de dados, por meio de atribuição de pesos às percepções e expectativas de qualidade dos especialistas. Na segunda etapa, a autora buscou avaliar de que forma o banco



de dados atendia a essas expectativas, por meio de um relacionamento anteriormente definido entre as características de qualidade e os tipos de não conformidades que ocorreram durante a utilização do sistema. Por último, a terceira etapa apresenta o acompanhamento da qualidade dos dados do sistema, por meio de gráficos e relatórios. Além disso, a autora desenvolveu e utilizou um protótipo AQUA, para validar o modelo e demonstrar a sua utilização.

Pinho (2001) explica como uma avaliação qualitativa de banco de dados, baseada na visão do usuário, pode ser feita. De acordo com a autora, inicialmente é fornecida para os avaliadores uma lista inicial de características, para que eles selecionem as características consideradas relevantes para a organização na ocasião da avaliação. Tais características recebem um parâmetro, que representa o seu grau de importância para o usuário, que pode variar de 0 a 4. Assim, tomando como base as características selecionadas pelo avaliador, pode-se prosseguir com o processo da avaliação, verificando como o sistema atende às expectativas do usuário. O processo descrito possibilita identificar possíveis deficiências dos dados, bem como a sua qualidade atual.

Para a avaliação baseada em não conformidades, é necessário que todos os tipos de erros possíveis de serem encontrados estejam bem estabelecidos, para que o usuário consiga identificar as não conformidades em um banco de dados. Em seu estudo, Pinho (2001) apresenta uma lista com uma descrição resumida de cada problema que pode ser encontrado e que afeta alguma característica de qualidade de dados.

Foi criado por Pinho (2001) um modelo matemático baseado em controle estatístico da qualidade, que preconiza a utilização de um Sistema de Deméritos para averiguar como as não conformidades podem auxiliar na melhoria do processo de produção. Assim, entendendo que nenhum produto é perfeito, definem-se categorias de gravidade da não conformidade. Em analogia ao banco de dados, cada vez que uma característica de qualidade era afetada por alguma não conformidade, o produto do banco de dados tem sua qualidade de alguma forma reduzida.

Com a utilização da ferramenta AQUA para realizar na prática a avaliação de dados pela não conformidade, Pinho (2001) verificou que essa pode ser uma ferramenta de análise de qualidade de dados útil, destacando, porém sugestões de melhoria como a automação do cadastro de erros ou não conformidades; a captação automática do tempo de utilização do produto avaliado e a associação a um sistema de controle de versões.

#### 4.2.2 Avaliação da qualidade em Data Warehouses

Amaral (2003) apresenta uma determinação de características de um ambiente computacional para ser utilizado em conjunto com o *Data Warehouse* no fornecimento de indicadores de qualidade dos dados ao usuário final. Assim, a autora apresenta o Ambiente de Suporte à Qualidade de Dados em *Data Warehouse* - AQUAWARE, que utiliza a abordagem de serviços *Web* para prover informações de qualidade para ferramentas de consultas analíticas, ou outras que atuem como clientes do *Data Warehouse*.

De acordo com a autora, o AQUAWARE tem como objetivo "possibilitar o fornecimento de informações sobre a qualidade de objetos do ambiente de *Data Warehouse*, pertencentes a qualquer nível da arquitetura de modelagem multicamadas". A autora ressalta que a possibilidade de associação entre as informações de qualidade e objetos de qualquer nível da arquitetura de modelagem multicamadas como um atributo importante do ambiente proposto, porque a qualidade das informações apresentadas está relacionada tanto com a qualidade dos dados quanto a objetos que pertencem aos demais níveis, como modelos, processos do *Warehouse* ou metamodelos. Assim, o AQUAWARE é formado por quatro itens que, ao serem incorporados ao ambiente de *Data Warehouse*, fornecem informações sobre a qualidade dos dados armazenados, que são: o Repositório de Metadados da Qualidade (RMQ), um conjunto de Cubos de Qualidade, uma Biblioteca de Componentes e um Serviço de Qualidade.

Amaral (2003) explica que a avaliação de qualidade do AQUAWARE segue o paradigma *Goal Question Metric* - GQM, proposto por Basili *et al.* (1994 *apud* AMARAL, 2003), no qual os parâmetros de avaliação são expressos em termos de objetivos, questionamentos e métricas, que ficam armazenados no Repositório de Metadados da Qualidade.

Os objetivos representam as necessidades apresentadas pelo cliente e cada um representa um propósito para um objeto do *Data Warehouse*, sob uma determinada perspectiva. As medidas de qualidade expressam o quanto o objetivo está sendo atendido e representam as respostas aos questionamentos, obtidas por meio das métricas, sendo geradas pelos componentes incluídos na Biblioteca de Componentes. A autora explica ainda que outros componentes da Biblioteca são utilizados, após a definição das medidas de qualidade, para calcular o grau de adequação aos objetivos propostos, para cada objeto do *Data Warehouse*, sendo, então, armazenadas nos Cubos de Qualidade, que são a fonte de consulta sobre a qualidade de dados do *Data Warehouse*. Já o Serviço de Qualidade

disponibiliza as informações de qualidade para os clientes, intermediando a relação entre o cliente e o serviço de consulta.

O trabalho de Amaral (2003) viabiliza o fornecimento de informações sobre a qualidade dos dados do *Data Warehouse* para ferramentas de consultas analíticas, para serem apresentadas ao usuário final, possibilitando que esse avalie a confiabilidade dos dados para a tomada de decisão. Além disso, o trabalho apresenta definições do XMLA-Q, uma versão adaptada do XMLA (*XML for Analysis*) (MICROSOFT, 2002) para abarcar as funcionalidades relativas à qualidade, a definição de indicadores para a avaliação da qualidade de dados sob a forma de objetivos, questionamentos e métricas, voltados para o usuário final. A autora apresenta sugestões para a melhoria da proposta, a exemplo da implementação das funcionalidades necessárias para a interação com o AQUAWARE em uma ferramenta de consulta analítica, para que essa possa receber informações de qualidade e apresentá-las ao usuário final, juntamente com os resultados das consultas; a definição de indicadores de qualidade para outros objetos do *Data Warehouse*, além das dimensões e dos fatos, bem como a especificação de componentes para a medição da qualidade desses e, a definição de indicadores de qualidade sob diferentes perspectivas para que os resultados da avaliação da qualidade sejam utilizados para aprimorar o processo de administração do ambiente de *Data Warehouse*.

#### **4.2.3 Avaliação da qualidade em Ambientes Transacionais**

Torelli (2005) apresenta um estudo no qual utiliza a metodologia da TDQM, associando a informação a um produto, para administrar essa informação em uma organização que apresentava perspectivas das administrações convencionais, adaptando a TQM em técnicas para um programa de QI. Assim, o autor apresentou estudos de caso para ilustrar como as empresas lidam com a questão da qualidade da informação, e para institucionalizar os programas de QI, bem como os sistemas de soluções em suas organizações. O autor aponta a dificuldade da institucionalização de programas de QI nas organizações, porque alguns paradigmas ainda são bem estabelecidos e difíceis de serem quebrados, porém, ele afirma que essa dificuldade tende a diminuir, com o aumento da disseminação da importância da qualidade dos dados em sistemas de informação.

Favaretto (2005) buscou realizar uma medição de Qualidade de Informação (o autor considera qualidade de informação e qualidade de dados como sinônimos) por meio de atribuição de valores para as dimensões de qualidade, de acordo com o ponto de vista do

usuário do produto. O autor sugere que as medições sejam realizadas de maneira automatizada, por meio de aplicativos específicos, mas, afirma que algumas medições podem também ocorrer de maneira não automatizada. A medição da QI é feita nos conjuntos de dados, decompostos em conjuntos menores, definindo-se previamente quais dimensões de qualidade são relevantes para os conjuntos, para que sejam medidas. De acordo com o autor, as dimensões são armazenadas em uma nova entidade do modelo de dados, que se relaciona com a entidade original que registra o dado em evidência, relacionando-se com uma nova entidade que registra os valores para as dimensões de QI a serem medidas.

Favaretto (2005) acrescenta ainda que a atribuição de valor para as dimensões pode ser feita automaticamente, quando os parâmetros a serem analisados no sistema de informação podem ser identificados. Já em situações em que o valor dependa de uma avaliação do usuário, essa atribuição não poderá, portanto, ser feita automaticamente.

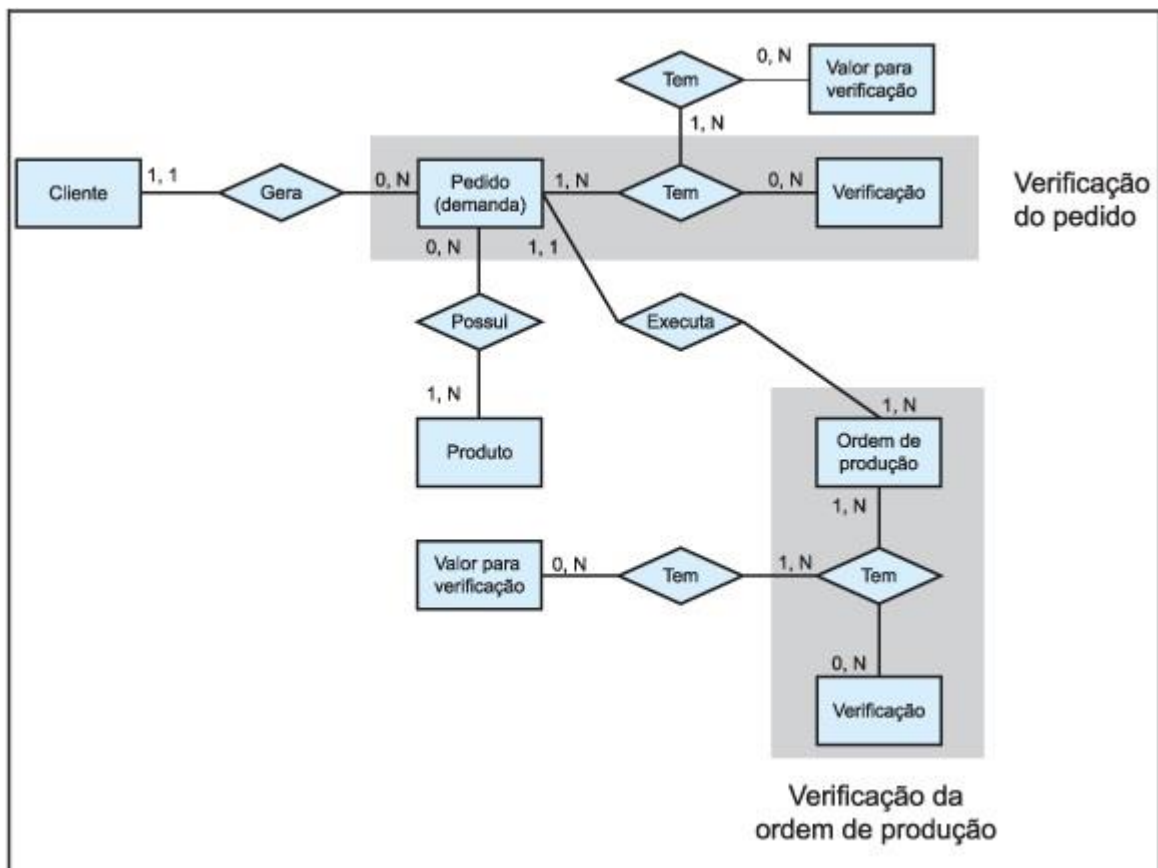
Em Favaretto (2007) é apresentado outro estudo, no qual se avaliou a relação entre o esforço despendido para a implantação da medição da qualidade de informação (QI) e os resultados obtidos, uma vez que a implantação da medição da qualidade da informação demanda tempo e recursos. O autor propôs uma metodologia para a medição da QI e apresenta os resultados de um experimento no qual a metodologia foi avaliada. A proposta desse trabalho é aplicada em uma das etapas do TDQM, que é a medição da qualidade de informação.

A aplicação da metodologia proposta seguiu uma série de etapas. Inicialmente escolheu-se para medição, o elemento do Sistema de Manufatura da Informação (SMI); também foram identificados os conjuntos de dados mais importantes, do ponto de vista do usuário. A etapa seguinte teve como objetivo medir a qualidade do SMI. Para essa etapa foi criada uma nova entidade no Modelo Entidade-Relacionamento, cujos registros contêm valores para o atendimento ou não da exigência do SMI.

Favaretto (2007) ressalta que a medição da qualidade do SMI foi feita com a adição de novas entidades e relacionamentos ao ME-R – *Modelo Entidade-Relacionamento* – do sistema em análise, permitindo avaliar se o SMI atende ao seu propósito, fazendo com que tanto o usuário do SMI quanto a equipe responsável pelo seu desenvolvimento e manutenção fiquem cientes quando ocorre uma não conformidade. A última etapa consistiu da medição da qualidade de informação, sendo consideradas como mais relevantes para a análise das dimensões Acuracidade, Credibilidade e Integridade. Nessa etapa, a entidade *Dimensão de qualidade de dados* foi incluída, bem como a entidade *Medida da qualidade de dados*. De acordo com o autor, a implantação dessas novas entidades e relacionamentos no sistema

original permitiu o registro de aspectos do monitoramento da qualidade e da informação gerada. É importante ressaltar que a medição das dimensões Acuracidade e Integridade pode ser feita automaticamente, mas a dimensão Credibilidade exige que a medição seja feita pelo usuário. A Figura 3 ilustra o ME-R com as modificações para medição da qualidade da SMI e da qualidade de informação. Para verificação da relação do esforço para a implantação da proposta e os resultados obtidos foi aplicado um questionário aos participantes da pesquisa.

**Figura 3 - Esquema do Modelo Entidade Relacionamento com as modificações necessárias para medição da Qualidade da SMI e da informação**



Fonte: Favaretto (2007).

Favaretto (2007) concluiu que para a medição e controle da qualidade da informação é necessário que uma quantidade significativa de esforço seja aplicada, contudo os resultados do experimento, ainda que preliminares, não demonstraram uma correlação significativa entre o esforço necessário para a implantação da medição da qualidade de informação e os resultados. O autor observa que esse resultado pode ter ocorrido porque os participantes do experimento não perceberam de forma clara relação entre o esforço e o resultado percebido, o que ocorre quando em situações de excesso de informação, o tomador de decisão tende a se

concentrar em informações centrais, deixando de lado informações "extras" no processo de tomada de decisão.

Favaretto & Vieira (2007) mediram a qualidade da informação utilizada e gerada no planejamento da produção por um conjunto de empresas da Região Metropolitana de Curitiba. Para alcançar o objetivo proposto, os autores elaboraram dois questionários que foram respondidos por profissionais da área de engenharia e administração, que atuam em empresas de produção, além de alunos de um curso de pós-graduação *lato sensu*, todos consumidores e geradores de informação.

Os questionários foram aplicados para verificar a qualidade das informações para duas etapas do planejamento da produção, a etapa de planejamento mestre e a etapa de planejamento detalhado, também conhecida como MRP – *Materials Requirements Planning*. A QI foi analisada nas dimensões Confiabilidade, Precisão e Temporalidade, com questões específicas sobre cada dimensão, que eram respondidas de acordo com uma escala variante de *péssima* (valor 1) a *excelente* (valor 5). Além disso, os autores explicam que no questionário sobre o planejamento da produção, foram feitas questões para cada dimensão das seguintes informações: previsão da demanda (independente) demanda dependente, pedidos em carteira e posição (nível) de estoque. Cada respondente indicou em uma escala de *inexistente* (valor 1) a  *muito freqüente* (valor 5) a sua participação na elaboração do plano mestre, e em uma escala de *inexistente* (valor 1) a  *muito grande* (valor 5) quanto à dependência das suas atividades em relação ao plano mestre. No questionário específico sobre o MRP, foram elaboradas questões para as dimensões das informações estrutura do produto, tempos de obtenção (*lead-times*) e posição de estoque, quando cada respondente indicou em uma escala de *inexistente* (valor 1) a  *muito freqüente* (valor 5) a sua participação na elaboração do MRP, e em uma escala de *inexistente* (valor 1) a  *muito grande* (valor 5) a dependência das suas atividades em relação ao MRP.

Como resultado, os autores observaram que as dimensões da QI tiveram conceitos entre médio e bom, enquanto para o MRP, as médias variaram entre os conceitos de *médio* e *bom*. Para os valores globais de todas as respostas para cada característica medida, observou-se que a QI percebida pelos participantes era pouco maior que a média. Dos resultados obtidos, os autores sugerem que a falta de participação dos entrevistados na elaboração do MRP pode ter sido o fator que levou a uma maior variabilidade na percepção da qualidade da informação. Ademais, os autores acrescentam que as informações utilizadas pelos respondentes dos questionários podem ser melhoradas, o que resultaria em melhora na qualidade do planejamento da produção e consequente o desempenho da empresa.

### 4.3 AVALIAÇÃO DA QUALIDADE SOB A PERSPECTIVA DO USUÁRIO

Calazans & Costa (2009) avaliaram a qualidade da informação por meio da identificação de aspectos de gestão que influenciam a informação, realizando estudos de caso em duas organizações bancárias federais brasileiras, construindo, posteriormente, um modelo para avaliação da qualidade da informação estratégica. Os autores utilizaram abordagens qualitativa e quantitativa, além da triangulação dos dados. De acordo com os autores, o modelo teórico conceitual da pesquisa estabelece que a avaliação da qualidade da informação estratégica envolve a classificação, avaliação da informação, bem como a identificação dos aspectos que afetam a qualidade. Tais características podem ser verificadas com a análise do processo de gestão da informação. Foram selecionados conceitos e modelos teóricos como base para a construção de um novo modelo conceitual, apontando as questões necessárias para avaliar a qualidade da informação estratégica no contexto das organizações bancárias. Para a construção do modelo alguns aspectos foram considerados, a saber: a classificação da informação quanto à dimensão, objetivos, fontes e abrangência; a avaliação, realizada com perspectiva do usuário; e a identificação dos aspectos que impactam a qualidade, através da análise da gestão da informação com a participação de diferentes interessados, e.g. produtores, custodiantes e consumidores/usuários.

O estudo tomou como base as percepções de profissionais que utilizavam a informação, ou, de alguma forma, participassem do processo de gestão. Assim, foram realizadas entrevistas semiestruturadas, questionários e pesquisa em documentos organizacionais, para as entrevistas utilizou-se amostra não probabilística intencional, baseado em critérios como experiência do entrevistado com informação estratégica e a função gerencial. A análise dos dados coletados foi feita através do teste do Qui Quadrado e de mineração de dados (Algoritmos Apriori) (STEVENSON, 2001 apud CALAZANS; COSTA, 2009), enquanto que para as entrevistas utilizou-se a técnica de análise de conteúdo e a interpretação envolveu a condensação de significados para identificar características comuns na percepção dos participantes (KVALE ,1996 apud CALAZANS; COSTA, 2009).

As características de qualidade foram avaliadas com relação à percepção do grau de importância identificados pelos usuários. Além disso, as características "livre de erro" e "fidedigna" foram apontadas como as mais relevantes e a percepção do grau de importância de cada característica apresentou padrões diferenciados para as duas instituições, de acordo com o objetivo e produto da área que trabalha com informação estratégica. Calazans & Costa

(2009) perceberam o resultado da análise dos dados coletados consideravam todas as fases do processo de gestão como importantes, contudo, "distribuição/disseminação da informação no momento, local e forma adequados" foi considerada a mais significativa para a qualidade da informação estratégica nas duas instituições estudadas. Ademais, a fase "tratamento/transformação da informação" foi mais significativa apenas em uma das instituições e a fase menos significativa para as duas instituições foi "definição de como e onde guardar a informação".

Costa & Calazans (2009) consideraram que o nível da qualidade da informação estratégica foi satisfatório e os procedimentos metodológicos utilizados credenciaram positivamente os resultados, já que o modelo construído e implementado demonstrou-se consistente e aplicável. Contudo, os autores asseveram que o modelo pode ser válido para casos similares ao do estudo. Como sugestões futuras são citadas:

(1) identificar, por meio da criação de uma ontologia, os conceitos e relações que existem no domínio escolhido ou contexto de interesse, para que os usuários tenham o mesmo entendimento sobre as características de qualidade que estão sendo avaliadas, e que a pesquisa seja realizada considerando os mesmos significados; (2) mapear as informações relevantes dentro do domínio estudado do contexto organizacional, de forma a priorizar a qualidade deste escopo de informações, garantindo uma ação mais efetiva para melhoria da qualidade; e (3) buscar a completude da pesquisa, não somente definindo ou adaptando um modelo, mas aplicando-o no mundo real, mapeando a percepção de qualidade e buscando identificar no processo informacional organizacional os aspectos que influenciam positiva ou negativamente essa percepção de qualidade.

Souza (2009) avaliou se a forma de apresentação dos indicadores de desempenho interferem na percepção de qualidade da informação pelos usuários da informação, na tentativa de indicar um melhor modelo para apresentação dos indicadores. Além disso, o autor buscou verificar se havia alguma relação entre a qualidade da informação e os processos desenvolvidos pelos consumidores da informação. Para o desenvolvimento do estudo foi desenvolvido um questionário para avaliar os aspectos relacionados à qualidade da informação nos indicadores de desempenho, para aplicar uma entrevista entre os funcionários de uma organização. Na entrevista foram avaliados um modelo de indicadores anteriormente utilizado pela empresa e um modelo de indicadores desenvolvido pelo autor. As pessoas entrevistadas eram funcionários da empresa relacionados à gestão da cadeia de suprimentos, no processo do pedido até a entrega do produto, sendo que alguns dos entrevistados tinham



poder de decisão ou alteração de diretrizes e ações, de acordo com o resultado dos indicadores, necessitando, portanto, diretamente de qualidade na informação recebida.

Baseado nas informações obtidas com a empresa, Souza (2009) escolheu cinco indicadores para avaliação: qualidade de fornecedores, qualidade para com clientes, desempenho de entrega de fornecedores, desempenho de entrega para com os clientes e custo operacional por unidade produzida. Além disso, o autor tomou por base o trabalho de Pipino *et al.* (2002), e elencou como dimensões mais importantes para o processo a acessibilidade, credibilidade, completeza, representação concisa, liberdade de erros, interpretabilidade e valor adicionado.

A partir do estudo foi possível observar que as dimensões acessibilidade, credibilidade e interpretabilidade apresentaram um bom nível de qualidade quando analisadas individualmente, enquanto as dimensões completeza, representação concisa, liberdade de erro e valor adicionado apresentaram um nível regular de qualidade. Ao analisar a combinação entre qualidade da informação do modelo de indicadores, observou-se que o modelo antigo apresentou um nível regular de qualidade.

A segunda parte do questionário buscou avaliar a relação entre a qualidade da informação e o processo de cada usuário, tentando identificar possíveis impactos do nível de qualidade em cada usuário. A análise dos questionários demonstrou que a qualidade da informação apresenta influência direta nos processos dos usuários.

Souza (2009) apresenta ainda um modelo de indicadores utilizado pela empresa, baseado na metodologia PCDA (*plan, do, check e act*), que apresenta o planejamento com métricas e metas, os gráficos dos resultados, a checagem das tendências e o plano de ação para melhoria dos resultados. Após a apresentação do novo modelo de indicadores, um novo questionário foi aplicado com os mesmos usuário que responderam à pesquisa anterior. Na segunda avaliação, a dimensão credibilidade apresentou um nível bom de qualidade da informação, enquanto as dimensões acessibilidade, completeza, representação concisa, liberdade de erros, interpretabilidade e valor adicionado apresentaram nível muito bom de qualidade, quando analisadas individualmente. Quando a combinação das dimensões da qualidade de informação do modelo de indicadores foi analisada, atingiu-se o nível muito bom de qualidade. Para a análise da relação entre a qualidade da informação e o processo de cada usuário, observou-se que com a melhora da qualidade da informação, os novos níveis de qualidade tiveram um pequeno acréscimo em seus processos, e a percepção para cada dimensão de qualidade é bastante similar entre os usuários.

Com esse estudo, Souza (2009) concluiu que a apresentação dos indicadores tem interferência na qualidade da informação observada pelos usuários, e a ocorrência de forte relação entre a qualidade das informações contidas nos indicadores e os processos desenvolvidos pelos colaboradores da empresa, facilitando a interpretação dos resultados de cada indicador e, conseqüentemente, o processo de tomada de decisão e o incremento de vantagem competitiva. Como recomendações, o autor sugere ampliar o estudo feito na empresa para outras dimensões de qualidade, incentivar o conhecimento dos conceitos e aplicações de qualidade da informação, bem como propor uma metodologia de medição da qualidade, de acordo com o perfil da empresa.

Em um estudo de caso, Kossowski & Favaretto (2010) buscaram explorar e operacionalizar a medição da qualidade de informação em um sistema integrado, de CRM (Sistemas de Gerenciamento de Relacionamento com o Cliente) e ERP (*Enterprise Resources Planning*), para identificar possíveis falhas relacionadas com a qualidade da informação, ocorrentes durante a integração entre sistemas. Ao final do estudo, os autores visavam também definir um método de análise de QI em sistemas integrados.

No estudo de caso foram pesquisados 89 funcionários de uma empresa de grande porte que utiliza a informação integrada do SAP/CRM, cuja linha de negócio era dividida em cinco processos críticos: (1) cadastro de clientes e produtos; (2) ingresso de pedidos; (3) faturamento; (4) contas a receber e (5) aplicação de pagamentos. Assim, os usuários entrevistados avaliaram a existência e importância das dimensões de qualidade, de acordo com a escala *Likert* (BOVEE, 2004 *apud* KOSSOWSKI; FAVARETTO, 2010), que apresenta sete níveis, variando desde "discordo plenamente" até "concordo plenamente". Esse pré-teste foi aplicado com vistas a ajustar a verificação do entendimento da pesquisa desenvolvida.

Com a aplicação do questionário, observou-se que a "Acessibilidade" foi a categoria mais importante e presente no sistema integrado. Em contrapartida, a categoria "Representação" demonstrou ser a de menor importância, o que acarretou em sugerir que a empresa avaliasse novamente esta categoria, para proporcionar maior satisfação ao negócio. A dimensão "Segurança no acesso" foi a de maior existência, enquanto "Temporalidade" e "Facilidade de entendimento" foram as dimensões mais pobres para os usuários de operacional da empresa. Já os analistas TI consideraram a "Objetividade" e "Acessibilidade" como dimensões mais pobres. Quanto à importância das dimensões, os autores observaram que os analistas de TI consideraram a "Representação" como dimensão de menor importância, enquanto os usuários da linha de negócio avaliaram "Valor agregado" como dimensão de

menor importância. Com os resultados obtidos, os autores sugeriram aos gestores da empresa que existisse um intercâmbio entre as áreas de TI e o negócio, visando o equilíbrio entre a necessidade de determinadas categorias e as dimensões de qualidade.

Kossowski & Favaretto (2010) concluíram que na empresa em estudo, havia divergência sobre os pontos de vista dos usuários do sistema e dos analistas de TI em relação ao processo de "Order to Cash". Tal divergência poderia ser justificada pelo fato de que enquanto os usuários finais enfrentam primeiramente os problemas de qualidade, os analistas trabalham com postura reativa, para corrigir o problema. Nesse sentido, os autores sugerem a criação de um grupo dedicado ao suporte e melhoria das ferramentas de trabalho, buscando, assim, um estreitamento de relações entre suporte e negócio. Finalmente, os autores reforçam a importância de maior integração entre usuários e analistas, principalmente no que se refere a uma atitude proativa por parte dos analistas, promovendo, dessa forma, maior eficiência ao negócio e melhor integração de dados. Novamente percebe-se que a avaliação da qualidade da informação foi feita de forma subjetiva, sem nenhum método automatizado.

Observa-se que, em nenhum dos trabalhos citados anteriormente, é apresentado um software como ferramenta para auxiliar a medição dos dados em tem de execução das transações.

#### 4.4 FERRAMENTAS DE AVALIAÇÃO DE QUALIDADE DE DADOS

Silva e Loh (2011) propõem uma ferramenta para auxiliar na tomada de decisão de fontes de informação em saúde na web, para que o público leigo consiga avaliar a credibilidade dessas fontes.

Segundo os autores, a avaliação automatizada da qualidade da informação na Internet não objetiva substituir a avaliação humana e subjetiva, mas sim auxiliá-la, uma vez que a avaliação humana é importante para definição de critérios. Ainda segundo Silva & Loh (2011), as maiores dificuldades encontradas na avaliação automatizada são os próprios critérios subjetivos que a página web pode apresentar, a exemplo de justificativas, ponto de vista, comprovação de evidência dentre outros, que necessitariam de implementação de algoritmos mais complexos.

Nota-se nesse trabalho uma tendência pela análise qualitativa, baseada na opinião de usuários ou pessoas envolvidas em quaisquer etapas da coleta ou processamento de dados ou

informações, notadamente nos trabalhos apresentados por Favaretto e autores colaboradores, que, comumente são realizados por meio de avaliação de questionários e atribuição de valores às respostas coletadas. Contudo, como Favaretto (2005) aponta, existem medições que requerem soluções específicas, em uma avaliação automatizada, como é possível observar com a utilização da ferramenta AQUA proposta por Pinho (2001) e AQUAWARE, por Amaral (2003). Além disso, é consenso entre os autores analisados que a qualidade dos dados/informação é fator imprescindível na tomada de decisões por parte dos usuários da informação.

#### 4.5 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentados alguns trabalhos necessários para o desenvolvimento desta dissertação. Foram discutidas algumas metodologias desenvolvidas para a avaliação da qualidade dos dados, como o desenvolvimento feito por Kovac *et al* (1997) de um programa de TDQM como incentivo da aplicação da gestão da qualidade em qualquer instancia empresarial, além da metodologia para a avaliação da qualidade da informação, chamada AIMQ, proposta por Lee *et al* (2001) e o modelo de qualidade de dados baseado em não conformidades apresentado por Pinho (2001), compatível com a ISO/IEC 9126. Foram discutidas também, uma proposta para a medição da qualidade dos dados em *Data Warehouses* (AMARAL, 2003) e algumas propostas de aplicação da TDQM em ambientes transacionais, como apresentado por Favaretto (2007) e Silva e Loh (2011).

Foi possível observar a importância da perspectiva do usuário na definição da qualidade dos dados, como pode ser visto no trabalho apresentado por Calazans & Costa (2009), que faz uso das percepções dos profissionais que utilizam a informação para a identificação de questões congruentes acerca de qualidade de dados, que nortearam a pesquisa. Além desse trabalho, Souza (2009) avaliou se a forma de apresentação dos indicadores de desempenho interferem na percepção de qualidade da informação pelos usuários da informação e Kossowski & Favaretto (2010) realizaram uma pesquisa em um grupo de funcionários de uma empresa de grande porte para avaliar a existência e importância das dimensões de qualidade de dados em um sistema integrado.

Observa-se também que a TDQM, bem como a definição de dimensões de qualidade para a medição da qualidade de dados, como proposto por Wand & Wang (1996), estão presentes em quase todos os trabalhos apresentados, o que demonstra a sua importância no que tange às questões de determinação da qualidade da informação.

A Tabela 4 ilustra as principais características encontradas nesses trabalhos para a avaliação da qualidade dos dados e que irão nortear o desenvolvimento desta dissertação.

**Tabela 4 - Características para a avaliação de qualidade de dados**

<b>Trabalhos</b>	<b>Avaliação baseada em não conformidades</b>	<b>Medição da qualidade dos dados</b>	<b>Avaliação com base na perspectiva do usuário</b>	<b>Avaliação da apresentação dos indicadores nos resultados</b>	<b>Avaliação Automatizada, em tempo de transações, dos Dados</b>	<b>Definição de dimensões de qualidade</b>
Amaral (2003)	Não	Sim	Sim	Não	Não	Sim
Calazans & Costa (2009)	Não	Não	Sim	Não	Não	Sim
Favaretto (2007)	Não	Sim	Sim	Não	Não	Sim
Kossowski & Favaretto (2010)	Não	Não	Sim	Não	Não	Sim
Kovac <i>et al</i> (1997)	Não	Sim	Sim	Não	Não	Sim
Lee <i>et al</i> (2001)	Não	Sim	Sim	Não	Não	Sim
Pinho(2001)	Sim	Sim	Sim	Não	Não	Sim
Silva e Loh (2011)	Não	Sim	Sim	Não	Não	Sim
Souza (2009)	Não	Sim	Sim	Sim	Não	Sim

Fonte: O Autor

Esta dissertação apresenta um modelo de avaliação de qualidade de dados baseado na proposta da TDQM, sendo utilizada como base, a definição de dimensões de qualidade de dados sob a ótica do usuário, cuja percepção de qualidade servirá para nortear a metodologia de determinação dessas dimensões. Para isso é proposto um processo de avaliação da qualidade dos dados e uma ferramenta que automatiza esse processo em algumas de suas etapas, ambas as propostas estão inseridas em um ambiente de avaliação a ser inserido no contexto da organização que terá seus sistemas avaliados e que é discutido no próximo capítulo.

## **CAPITULO 5 – UM AMBIENTE PARA A AVALIAÇÃO DA QUALIDADE DE DADOS EM AMBIENTES TRANSACIONAIS**

Neste capítulo será apresentado um modelo de processo para a medição de qualidade de dados baseado nas etapas da TDQM, utilizando uma ferramenta proposta neste trabalho. Este modelo deverá ser instanciado para cada sistema a ser avaliado, como será demonstrado no Capítulo 6.

A ferramenta proposta nesta dissertação é denominada FEMEQ - Ferramenta para Medição da Qualidade em Bancos de Dados Transacionais e a sua principal funcionalidade é permitir que o usuário de um sistema de informação possa monitorar os dados gerados pelo sistema em análise, quantificando o nível de qualidade dos dados, através de um sistema de métricas definidos por ele, utilizando cálculos básicos da estatística descritiva, a fim de permitir a análise dos resultados de uma maneira rápida e eficiente.

Na Seção 5.1 é apresentada a FEMEQ e os aspectos relacionados à sua implementação. Na Seção 5.2 é apresentado o modelo de processo para a avaliação de dados baseado na perspectiva do usuário do sistema de informação em análise e na FEMEQ. Por fim, na Seção 5.3 são discutidas as considerações finais a respeito deste ambiente para a avaliação da qualidade de dados em ambientes transacionais.

### **5.1 FEMEQ - FERRAMENTA PARA MEDIÇÃO DA QUALIDADE EM BANCOS DE DADOS TRANSACIONAIS**

Para o desenvolvimento da FEMEQ, serão apresentados nesta seção as premissas adotadas em seu desenvolvimento, os requisitos funcionais e não funcionais, a arquitetura da FEMEQ um modelo lógico de dados da ferramenta e as definições adotadas para a implementação da FEMEQ.

#### **5.1.1 Premissas para o Desenvolvimento da FEMEQ**

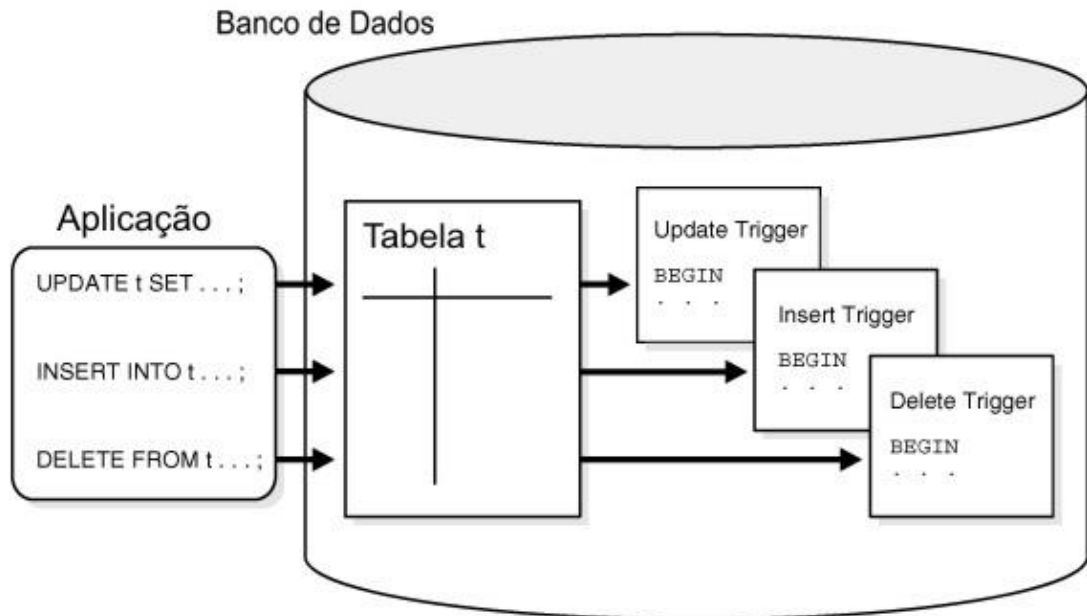
Como visto no Capítulo 3, Seção 3.3, foram discutidos aspectos acerca da TDQM, que fornecem a base para a maioria dos trabalhos discutidos no Capítulo 4. Dada a sua relevância, a abordagem por ela apresentada norteará a construção da ferramenta proposta. Além disso, o desenvolvimento da ferramenta foi baseado em procedimentos armazenados de banco de

dados para o monitoramento das transações ocorridas nas bases em avaliação. Para tanto, esses procedimentos são disparados a cada transação através de gatilhos, também conhecidos como *trigger*. *Trigger*, ou gatilho, segundo a Oracle® Database Concepts, é um bloco de programação para bancos de dados relacionais, escrito em linguagem PL/SQL – Procedural Language/Structured Query Language, associado a uma tabela específica, cujo código fica armazenado no próprio banco para que seja executado automaticamente sempre que um evento oriundo de uma instrução DML – Data Manipulation Language for aplicado nessa tabela. Para este trabalho, as principais vantagens do uso das *triggers* são devido ao fato que elas podem conter uma lógica de processamento complexa e serem executadas de maneira transparente para a aplicação cliente. Uma *trigger* pode ser definida para ser executada antes ou depois que ocorrerem um dos três tipos de ações possíveis: inserção, exclusão ou atualização dos dados de uma tabela, como mostra a Figura 4.

Na ferramenta proposta, o monitoramento da qualidade é realizando em tempo real, sendo os valores referentes à quantificação da qualidade dos dados armazenados na base de dados do sistema em avaliação a cada transação que utilize os dados que estão sendo avaliados.

Em sua solução, Amaral (2003) propõe a criação de um serviço de qualidade que funciona como uma camada entre o Data Warehouse e o sistema cliente, na qual é realizada a extração dos dados. No caso da FEMEQ, foi definido que a utilização de *triggers* para o monitoramento dos dados seria uma solução viável, já que a ferramenta não fica dependente de um serviço externo, comprometendo os resultados em razão de falhas do serviço, além de permitir um melhor aproveitamento dos recursos dos SGBDs e diminuir o esforço do usuário durante a avaliação dos dados. Sendo assim, o monitoramento dos dados ocorrerá em tempo real e estará ativo em todas as transações realizadas no banco de dados. A partir da análise dos trabalhos correlatos, discutidos no Capítulo 4, observou-se que não há proposta semelhante em nenhum deles, i.e. nenhum dos trabalhos apresentados propõe uma ferramenta automatizada que permita a avaliação da qualidade dos dados em tempo de execução das transações baseado nas , desta forma, esta é uma relevante contribuição deste trabalho. Este modelo de monitoramento será mais bem discutido na Seção 5.2.2 deste capítulo.

Figura 4 – Funcionamento de uma trigger



Fonte: Adaptado de Oracle® Database Concepts.

Com a finalidade de auxiliar o avaliador na criação das *triggers* de acordo com as regras de negócio relacionadas à avaliação da qualidade dos dados do sistema em avaliação, foi proposto que a FEMEQ forneça alguns *templates* para facilitar o seu desenvolvimento. Esses *templates* são modelos de código em PL/SQL, alteráveis, que nortearão a criação de *triggers* ficando a cargo do avaliador apenas adaptá-lo às necessidade de negócio do sistema em avaliação. Além disso, é permitido que o avaliador também cadastre os seus *templates* para posterior uso, já que a ferramenta possui essa funcionalidade.

Assim como os trabalhos propostos por Favaretto (2007) e por Amaral (2003), a FEMEQ funciona como uma extensão da base de dados do sistema em avaliação, como proposto por Storey (1998). E, para que isso seja possível são adicionadas ao modelo lógico da base de dados do sistema em avaliação algumas tabelas de dimensões de dados. Essa atividade será mais bem descrita na Seção 5.2.1.

A terceira fase da TDQM, i.e. a fase de Análise, propõe que os resultados obtidos no monitoramento dos dados sejam analisados para a certificação da Qualidade da Informação (QI). Para auxiliar nessa tarefa, a FEMEQ oferece relatórios utilizando algumas variáveis da estatística descritiva. Segundo Guedes *et al* (2012), o objetivo básico da estatística descritiva



é sintetizar um conjunto de valores de mesma natureza, possibilitando uma visão global de sua variação.

Na FEMEQ são utilizadas as medidas estatísticas *Número de Elementos, Média, Variância, Desvio Padrão e Coeficiente de Variação*. Guedes *et al* (2012) afirma que “se duas populações apresentam a mesma média, mas os desvios padrão não são iguais, isto não significa que as populações têm o mesmo comportamento”. O coeficiente de variação permite avaliar a homogeneidade do conjunto de dados, além de mostrar se a média representa os dados com clareza. Guedes *et al* (2012) afirma que “um coeficiente de variação superior a 50% sugere alta dispersão o que indica heterogeneidade dos dados”. O cálculo para essas medidas é explicado em detalhes na Seção 5.2.3.

O modelo MDI – Multiple Document Interface (BORLAND, 2002) é um conceito usado em aplicações com formulários do tipo janela que permite a abertura de uma gama de formulários organizados dentro de um formulário pai proporcionando ao usuário um maior conforto em relação ao modelo SDI – Single Document Interface (BORLAND, 2002), no qual o usuário é obrigado a trabalhar com diversas janelas distribuídas na tela, o que pode dificultar a sua visualização, ou ter que visualizar uma janela por vez. O modelo MDI utiliza o conceito de *container* que tem a capacidade de abrigar diversos outros formulários, sendo que o principal (container) é chamado de formulário pai e abriga os formulários filhos, aproveitando melhor o espaço da tela. A principal razão da escolha desse modelo é permitir uma melhor organização dos formulários existentes na FEMEQ, não possibilitando que o usuário manipule esses formulários fora da área de trabalho da ferramenta.

### **5.1.2 Requisitos Funcionais e Não Funcionais**

Os requisitos de um sistema servem para ajudar os engenheiros de software a entenderem de maneira mais aprofundada o problema que irão dispende esforços para resolver, entendendo qual o impacto que o software desenvolvido irá exercer sobre o negócio do cliente (PRESMAN, 2006). Em Sommerville (2007), os requisitos de um sistema são descritos como definições das restrições operacionais e das funções que devem ser fornecidas por esse sistema. Com base nessa afirmação, pode-se entender que os requisitos de software servem para compreender quais as funcionalidades e restrições um software a ser desenvolvido deve possuir, a fim de satisfazer ao máximo possível, a necessidade do cliente.

## ➤ Requisitos Funcionais

Somerville (2007) classifica os requisitos funcionais como sendo a definição das funcionalidades que ditam como o sistema deverá reagir em condições específicas e como se comportar em determinadas situações, bem como o que o sistema não deve fazer. Pressman (2006) chama esses requisitos de *requisitos normais* e, para ele, são aqueles que refletem os objetivos e metas definidos para um produto de software durante as reuniões com o cliente.

Para o desenvolvimento da FEMEQ, foram identificados, pelo autor, os seguintes requisitos funcionais:

- A ferramenta deverá permitir o monitoramento simultâneo de mais de um Sistema de Informação;
- Deverá ser disponibilizada a consulta aos projetos cadastrados;
- A ferramenta deverá permitir a seleção das dimensões de QI a serem monitoradas nos SI;
- Deverá ser permitida a adição de novas dimensões de QI na ferramenta para a análise da qualidade de dados;
- Deverá ser disponibilizada a consulta às dimensões cadastradas;
- Deverá ser flexível e aceitar diferentes métricas de qualidade;
- Poderão ser adicionados *templates* de scripts para a criação de *triggers*;
- Deverá ser disponibilizada consulta aos *templates* cadastrados;
- Os relatórios devem ser adicionados à ferramenta sem a interferência do usuário durante a seleção das dimensões a serem monitoradas;
- Os relatórios deverão ser gerados sob demanda;
- Os relatórios devem ser disponibilizados por dimensão testada para cada sistema;
- Deverá ser permitida a seleção da data de início e fim do monitoramento para seleção dos resultados na apresentação dos relatórios;
- A ferramenta deverá permitir o cadastramento de usuários para a sua utilização.

### ➤ Requisitos Não Funcionais

Os requisitos não funcionais são aqueles que não estão explícitos na definição dos produtos. Somerville (2007) apresenta esses requisitos como sendo as restrições sobre os serviços e funções oferecidas pelo sistema, tendo origem nas necessidades implícitas dos usuários. Pressman (2006) chama esses requisitos de *Requisitos Esperados* e os define como os requisitos implícitos que podem ser tão importantes que, por vezes, o cliente não consegue identificar, mas contribuem diretamente na sua satisfação.

Os requisitos não funcionais definidos para a FEMEQ são:

- Deverá ser compatível com o sistema operacional Windows;
- A ferramenta deverá apresentar um conjunto de dimensões pré-cadastradas para facilitar a sua utilização;
- Os dados deverão ser monitorados sem o auxílio de ferramenta externa;
- Ser baseada em uma arquitetura cliente/servidor;
- Permitir a utilização do padrão SQL3 para a programação de *triggers*;
- Somente os usuários cadastrados poderão ter acesso ao sistema e utilizar seus recursos;

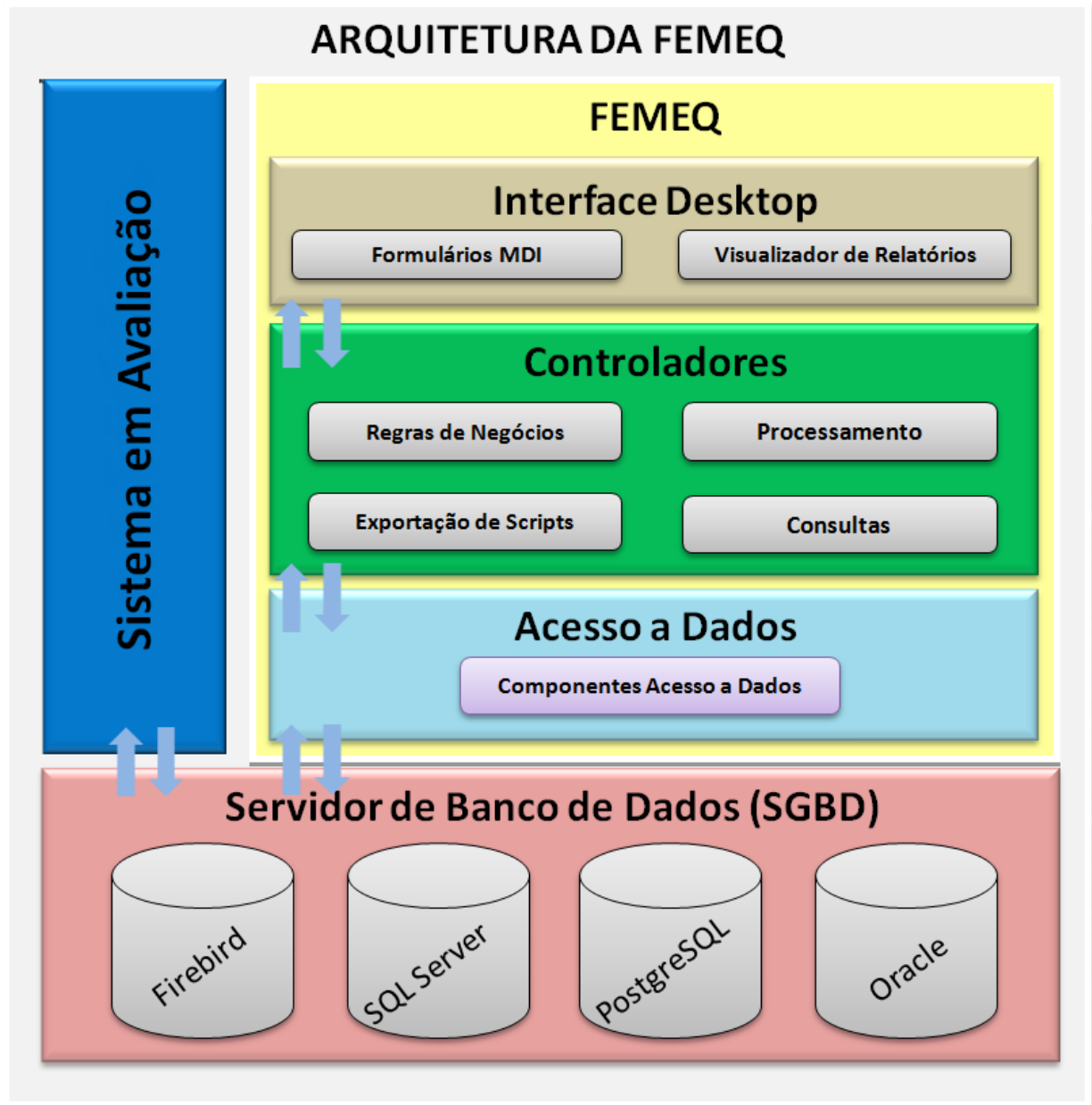
#### 5.1.3 Arquitetura da FEMEQ

A arquitetura proposta para a FEMEQ, apresentada na Figura 5, é dividida em 5 camadas: (i) *Interface Desktop*, camada em que se encontram os componentes responsáveis por realizar a interação do sistema com os usuários; (ii) a camada *Controladores*, na qual é realizado o processamento das regras de negócio da ferramenta; (iii) *Acesso a Dados*, camada responsável por realizar a persistência dos dados nas bases de dados, tanto da ferramenta, quanto dos sistemas em avaliação; (iv) *Servidor de Banco de Dados*, camada composta pelos SGBDs dos sistemas em avaliação, além da base de dados da ferramenta, sendo responsável por executar as requisições da camada *Acesso a Dados* e (v) *Sistema em Avaliação*, camada que realiza a integração do sistema que será monitorado com a FEMEQ.

O modelo de arquitetura proposto foi baseado na arquitetura MVC – Modelo, Controle e Visão ou *Model-view-control* (BURBECK, 1992), por permitir uma melhor organização do código de programação, em razão da separação entre as camadas *modelo*, na qual estão os componentes de persistência de dados e de negócio, a camada de *visualização*, onde se encontram os componentes responsáveis por fornecer a interface para a interação com o

usuário e a camada de *controle* que serve como mediadora entre as camadas de *visão* e *modelo*.

Figura 5 - Arquitetura da FEMEQ



Fonte: O Autor

Na Figura 5 é possível verificar que, tanto a camada *Acesso a Dados*, quanto *Sistema em Avaliação* requisitam os serviços do SGBD para a sua execução. O SGBD, então funciona como um integrador entre as camadas *Acesso a Dados* e *Sistema em Avaliação*, sendo essencial para o funcionamento da FEMEQ. Esta arquitetura possibilita a integração entre FEMEQ e o sistema em avaliação, uma vez que as tabelas que armazenam os valores resultantes do monitoramento dos dados em avaliação se encontram na base de dados do

sistema e são acessadas pela FEMEQ para a elaboração dos relatórios de qualidade. A arquitetura foi definida em camadas para permitir um baixo acoplamento entre os seus componentes, embora ocorra a integração da FEMEQ com o sistema em avaliação.

Cada camada se comunica diretamente com a camada superior, enviando e recebendo requisições e oferecendo seus serviços. As camadas possuem funções definidas que são descritas a seguir:

- (i) *Interface Desktop* é camada formada pelos componentes *Formulário MDI*, que contém as telas para a inserção de dados na FEMEQ e são responsáveis por desenhar essas telas, permitindo a interação da ferramenta com os usuários, apresentando os dados resultantes dos processamentos executados na camada *Controladores*, além de receber as requisições dos usuários. O componente *Visualizador de Relatórios* possui a finalidade de permitir que relatórios sejam criados e apresentados para a visualização dos usuários. As requisições dos usuários são recebidas por essa camada e enviadas para a camada *Controladores*;
- (ii) A camada *Controladores* é a que encapsula as classes e componentes responsáveis por realizar os processamentos requisitados pelo usuário através da camada de *Interface Desktop*. Esta camada é composta por quatro componentes, ilustrados na Figura 5, cujas funções são:
  - a. *Regras de Negócio* - contém as regras para persistência dos dados. Esse componente é responsável por realizar as verificações das regras de negócios definidas para o armazenamento e apresentação dos dados;
  - b. *Processamento* - realiza os processamentos inerentes à preparação dos *scripts* de banco de dados para a criação das tabelas de dimensões, tabelas de valores de dimensões e *triggers* de monitoramento, além dos cálculos para a geração dos relatórios e verificações de segurança;
  - c. *Exportação de Scripts* - após a execução dos processamentos para a preparação dos *scripts* de banco de dados, em uma linguagem de definição de dados e dos *scripts* de criação de *triggers*, em uma linguagem de programação de banco de dados, um conjunto de componentes são responsáveis pela exportação de *scripts*, i.e. o envio desses *scripts* para o sistema gerenciador de banco de dados do sistema em avaliação, representado na Figura 5 pela camada *Servidor de Banco de Dados*

(*SGBD*), para a criação das tabelas de dimensões, valores de dimensões e *triggers*;

- d. *Consultas* – componente responsável por realizar consultas aos dados contidos nas bases de dados, tanto da ferramenta, quanto do sistema em avaliação, a exemplo, as dimensões de avaliação e os valores resultantes dos monitoramentos. As consultas são preparadas por um conjunto de componentes pertencentes à camada *Controladores* e repassadas para a camada *Acesso a Dados*. Os resultados dessas consultas são, então, enviados por esse componente *Consultas* à camada *Interface Desktop* para a apresentação ao usuário;
- (iii) A camada *Acesso a Dados* é composta por um conjunto de componentes responsáveis realizarem a persistência dos dados nas bases de dados da ferramenta e do sistema em avaliação. As consultas preparadas na camada *Controladores* também são executadas pelos componentes da camada *Acesso a Dados* e os resultados dessas consultas enviados à camada *Controladores* para serem tratados pelos componentes *Regras de Negócio*, de acordo com as regras definidas para a formatação de apresentação e, posteriormente, apresentados na camada *Interface Desktop*;
- (iv) *Servidor de Banco de Dados* - a FEMEQ foi concebida para realizar a avaliação da qualidade de dados nos mais diversos tipos de servidores de banco de dados para garantir a sua generalidade. A camada *Servidor de Banco de Dados (SGBD)* é composta pelo sistema gerenciador de banco de dados que irá gerenciar o acesso aos dados do sistema em avaliação;
- (v) *Sistema em Avaliação* – essa camada é composta pelo sistema que será monitorado pela FEMEQ para a avaliação da qualidade dos dados por ele gerados e manipulados. A sua integração com a camada *Acesso a Dados* ocorre por intermédio da camada *Servidor de Banco de Dados*.

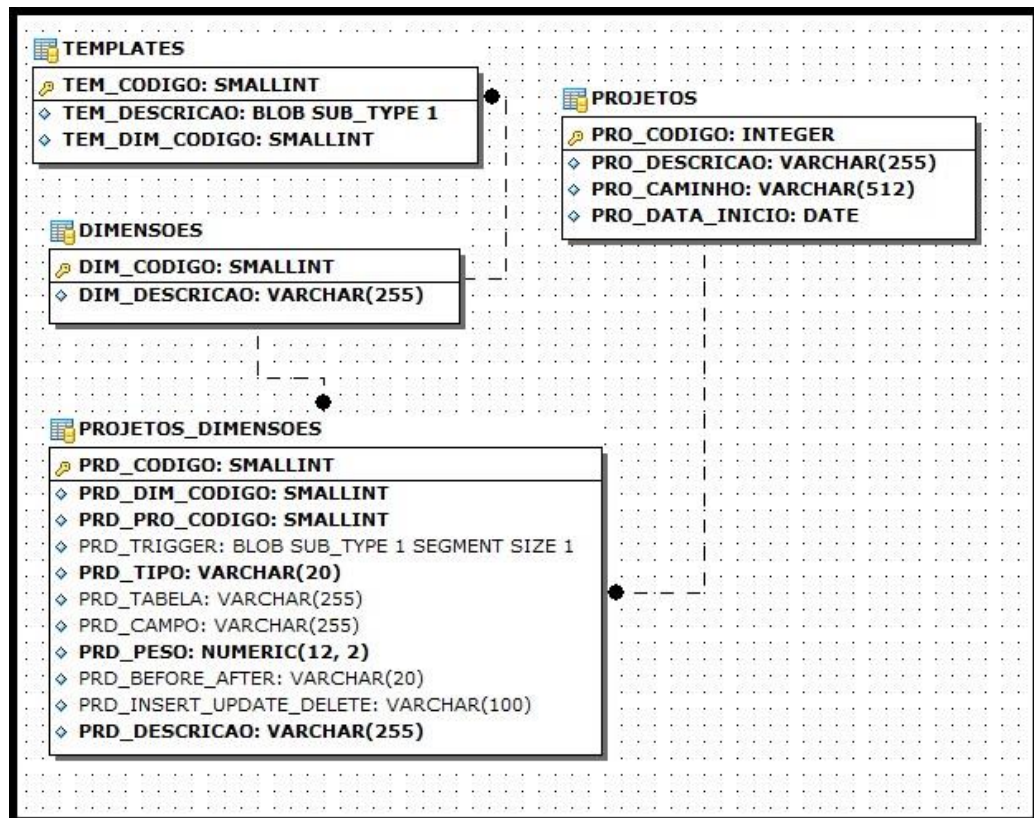
#### 5.1.4 O Modelo Lógico de Dados da FEMEQ

O *Modelo Lógico de Dados* é um esquema de banco de dados que descreve as estruturas contidas no banco, com uma representação lógica das informações da área de negócios do SI, em uma abstração relacional de dados com mais detalhes de implementação que o *Modelo Conceitual de Dados* (HEUSER, 2004).

O modelo lógico tem como finalidade descrever as estruturas que estarão contidas em um banco de dados sem considerar, ainda, nenhuma característica específica de um Sistema Gerenciador de Banco de Dados (SGBD), resultando em um esquema lógico de dados sob a ótica da abordagem relacional de dados. Nesta seção o modelo lógico de dados usado é apresentado e ilustra como a FEMEQ armazena os metadados de testes, i.e. as dimensões, métricas e *triggers* definidas pela análise do sistema em avaliação da qualidade. Esse modelo é ilustrado na Figura 6.

A FEMEQ foi concebida para realizar a avaliação de mais de um sistema por vez. Portanto, se faz necessário o cadastramento de todos os sistemas em avaliação na ferramenta. Para o armazenamento dos dados referentes a esses sistemas, é utilizada uma tabela chamada *Projetos*, que pode ser vista na Figura 6. As dimensões para a avaliação da qualidade de dados devem ser informadas à FEMEQ para sua utilização durante a avaliação dos sistemas. Estas dimensões são, então, armazenadas em uma tabela chamada *Dimensões*. A FEMEQ permite que sejam cadastrados diferentes modelos de código PL/SQL para a criação de *triggers* de monitoramento, sendo que cada modelo cadastrado pode estar relacionado a mais de uma dimensão. O armazenamento desses códigos é realizado na tabela *Templates*. Por fim, para cada sistema em avaliação, deve ser informada um conjunto de dimensões para o monitoramento e avaliação da qualidade, tendo a tabela *Projetos\_Dimensões* a finalidade de representar a agregação entre essas duas entidades.

Figura 6 - Modelo Lógico de Dados da Ferramenta



Fonte: O Autor

O dicionário de dados do modelo lógico de dados ilustrado na Figura 6 é apresentado no Apêndice A.

### 5.1.5 Implementação da FEMEQ

O desenvolvimento da FEMEQ foi realizado na ferramenta RAD (Rapid Application Development) Delphi, em sua versão 7, utilizando como base o paradigma de programação orientado a objetos, graças às suas propriedades, como reusabilidade de código, escalabilidade e manutenibilidade, o que permite uma programação mais dinâmica, com uma capacidade de evolução da ferramenta sem comprometer a sua integridade, utilizando a linguagem Object Pascal. Essa ferramenta foi escolhida em razão da sua maturidade. Tal ferramenta associa-se com facilidade à API (Application Programming Interface) do sistema operacional Windows, permitindo que os programas gerados por ela explorem ao máximo os recursos desse ambiente.

A principal vantagem da utilização do Delphi é que, segundo sua documentação, (BORLAND, 2002), sua utilização oferece dois níveis de programação: a programação em



*designer*, que utiliza recursos de programação visual, fazendo uso de componentes prontos, reduzindo o esforço do programador e fornecendo agilidade e facilidade à criação dos programas, além da redução dos erros, já que muitos desses componentes já foram testados exaustivamente e a programação *component writer*, no qual é possível se criar componentes para serem utilizados no desenvolvimento dos programas. Portanto, a escolha do Delphi em sua versão 7, se dá por sua estabilidade, facilidade de aquisição de componentes gratuitos e, apesar de ter sido distribuída no ano de 2002, ainda se apresenta compatível com as diversas distribuições do SO Windows existentes no mercado.

A linguagem Object Pascal, como afirma a documentação do Delphi, (BORLAND, 2002) é híbrida, já que é possível desenvolver software baseados no paradigma estruturado da linguagem Pascal, como no orientado a objetos – OO – e mostra-se bastante poderosa por possuir as principais características da programação OO, como classes e objetos, além de permitir a utilização de interfaces, assim como as linguagens Java e C#, tratamento de exceções e programação multithreaded.

A ferramenta FEMEQ foi concebida para ser utilizada em bases de dados que utilizam o SGBD – Sistema Gerenciador de Banco de Dados – Firebird 2.1, porém esforços já estão sendo despendidos para que sejam aceitos outros SGBDs, como o SQL-Server da Microsoft e Postgres. A sua escolha é justificada por ser um SGBD gratuito, confiável e que implementa a maioria das funcionalidades previstas no padrão SQL3.

Para a concepção da ferramenta, foram utilizadas como base as fases para a manutenção da qualidade definidos pela TDQM. Sendo assim, os requisitos foram definidos utilizando essa metodologia. O armazenamento dos dados referentes aos projetos em avaliação, modelos de *triggers* e dimensões é realizado em uma base de dados, cujo modelo relacional é apresentado na Seção 5.1.4 deste capítulo.








Para se realizar a avaliação dos dados com base na FEMEQ, se fez necessária a definição de um modelo de processo para a utilização da ferramenta, a qual será descrita na Seção 5.2. A utilização desse modelo permite que o avaliador defina com maior clareza os dados a serem avaliados, bem como as dimensões e métricas para a avaliação desses dados. A definição de métricas é importante porque possibilita a quantificação da qualidade, possibilitando ao avaliador uma análise baseada em dados estatísticos dos resultados do monitoramento da qualidade dos dados.

## 5.2 MODELO DE PROCESSO PARA AVALIAÇÃO DA QUALIDADE DOS DADOS BASEADO NA PERCEPÇÃO DOS USUÁRIOS

Nesta seção é apresentado um modelo de processo para a avaliação da qualidade de dados de sistemas de informação baseado na percepção dos usuários, para isso um diagrama de atividades referentes à avaliação da qualidade utilizando a FEMEQ foi elaborado para ilustrar as etapas do processo, além dele é proposta a extensão do modelo lógico de dados do sistema em avaliação com a adição das entidades referentes às dimensões de qualidade, e são discutidos o monitoramento dos dados, as variáveis que compõem a geração dos relatórios e uma proposta de modelo de processo para a definição das dimensões de qualidade sob a perspectiva do usuário.

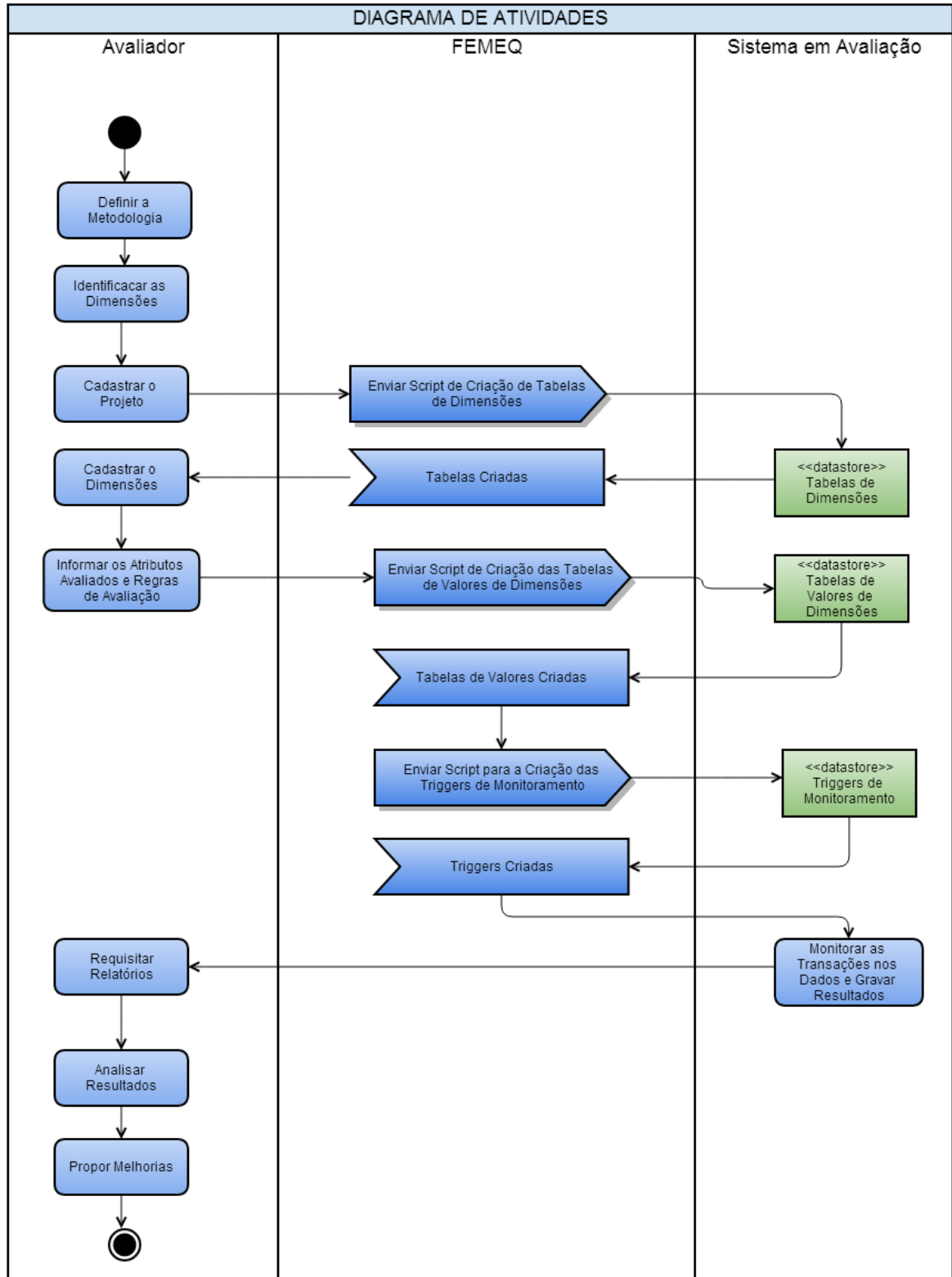
Para a avaliação de qualidade dos dados em um sistema de informação foi definido, neste trabalho, um processo de avaliação da qualidade dos dados baseado na percepção dos usuários. Este processo será descrito a seguir e é ilustrado graficamente por meio de um diagrama de atividades. O diagrama de atividade é um gráfico de fluxos, pertencente à linguagem UML – Unified Modeling Language – utilizado para modelar o aspecto comportamental dos processos, mostrando o fluxo de controle de uma atividade para a outra. A sua utilização facilita o entendimento das ações a serem executadas para que as atividades de um determinado processo sejam cumpridas. A Figura 8 mostra o diagrama de atividades que modela as atividades que devem ser executadas no processo de avaliação da qualidade dos dados de um SI utilizando a ferramenta FEMEQ. Para facilitar o entendimento do diagrama de atividades apresentado na Figura 8, é apresentada uma legenda explicativa na Figura 7.

**Figura 7 - Legenda de Símbolos UML**

Símbolo	Descrição
	Início do fluxo de atividades
	Atividade
	Sentido do fluxo
	Final do fluxo de atividades
	Objeto de Banco de Dados
	Ação de envio de sinais para um objeto ou ação
	Espera de ocorrência de um evento

Fonte: Adaptado de Guedes (2009)

**Figura 8 - Diagrama de Atividades para Avaliação de Qualidade Utilizando a FEMEQ**



Fonte: O Autor

A primeira atividade para a avaliação da qualidade é a definição de uma *metodologia* para o levantamento das dimensões de qualidade, definição de métricas para a quantificação do resultado do monitoramento dos dados para a avaliação da qualidade e regras de negócio para a medição de cada dimensão. A Seção 5.2.4 deste capítulo apresenta uma proposta de modelo para execução desta atividade. Aplicando esse modelo proposto, é possível *identificar as dimensões* de qualidade dos dados sob a perspectiva do cliente do SI em avaliação, determinando quais dados deverão ser analisados. É preciso, então, *cadastrar o projeto* na FEMEQ para que a ferramenta *envie o script de criação das tabelas de dimensões* para SGBD o sistema em avaliação, que é um conjunto de códigos em uma linguagem de definição de dados – DDL, que adicionam as *tabelas de dimensões* nessa base, a fim de que sejam armazenados os metadados, i.e. os dados referentes ao que vai ser avaliado, para a realização dos testes. Com as *tabelas de dimensões criadas*, deve-se *cadastrar as dimensões* que serão aplicadas e informar os *atributos avaliados e as regras de avaliação*, representadas por *triggers* que realizam, efetivamente, a avaliação dos dados. A FEMEQ, então, *envia o script de criação das tabelas de valores de dimensões* para o SGBD do SI em avaliação, que é um conjunto de códigos DDL e servem para adicionar as *tabelas de valores de dimensões* que, depois de confirmada a adição, representada no digrama pelo componente da figura que mostra o evento *tabelas de valores criadas*, irá *enviar script para a criação das triggers de monitoramento* para o SGBD do sistema em avaliação, i.e. *scripts* de banco de dados em linguagem PL/SQL para a criação *das triggers de monitoramento*. Após o evento *triggers criadas*, essas passam a *monitorar as transações nos dados e gravar resultados* nas tabelas de valores na base de dados do SI em avaliação. Essas transações são as operações de inserção, atualização e exclusão que acontecem com os dados e são continuamente avaliadas pelas *triggers* de monitoramento. Com os dados gerados, o avaliador pode então *requisitar relatórios* à interface da FEMEQ que servirão de base para que ele possa *analisar os resultados e propor melhorias*, realizando uma auditoria nos processos que estejam apresentando resultados negativos.

A seguir são descritos os elementos que compõem o processo ilustrado na Figura 8:

- **Avaliador** é a pessoa que executa a análise do SI em avaliação. Ele é responsável por identificar quais dimensões devem ser aplicadas e quais atributos na base de dados do cliente. Nesse processo, é importante que ele determine uma metodologia a fim de obter as dimensões que serão avaliadas para a medição da qualidade dos dados;

- **FEMEQ** é a ferramenta que desenvolve o papel de coletar os valores resultantes da medição da qualidade dos dados para a avaliação;
- **Sistema em Avaliação** é o sistema de informações que será avaliado com a ajuda da ferramenta FEMEQ;
- **Definir Metodologia** é a atividade em que o avaliador deve definir um modelo metodológico que lhe permita determinar quais são as necessidades de qualidade da informação para o SI avaliado sob a perspectiva do usuário desse SI;
- **Identificar as Dimensões** é a ação de identificar as dimensões de qualidade de dados de acordo com a metodologia definida pelo avaliador;
- **Cadastrar Projeto** ação de cadastrar o projeto, SI avaliado, na ferramenta para permitir o seu monitoramento;
- **Enviar Script de Criação das Tabelas de Dimensões** é o envio dos sinais em uma linguagem de definição de dados – DDL – para a criação das tabelas de monitoramento da qualidade como uma extensão modelo de dados do sistema em avaliação. A adição entidades de dimensão ao modelo lógico do sistema em avaliação será mais bem discutida na Seção 5.2.1;
- **Tabelas de Dimensões** são as tabelas que armazenam quais são as dimensões e os atributos que serão avaliados na base de dados do sistema em avaliação;
- **Tabelas Criadas** é uma resposta que o SGBD retorna à FEMEQ informando que as tabelas foram criadas na base de dados do SI em avaliação;
- **Cadastrar Dimensões** é o cadastramento das dimensões de avaliação;
- **Informar os Atributos Avaliados e Regras de Avaliação** é o cadastramento dos atributos a serem avaliados, informando quais as dimensões serão avaliadas para esses atributos, bem como as regras para a avaliação, fornecidas através de códigos PL/SQL – Procedural Language/Structured Query Language – para a criação das *triggers*;
- **Enviar Script para a Criação das Tabelas de Valores de Dimensões** é o envio dos sinais em uma linguagem de definição de dados para a criação das tabelas que armazenam os valores resultantes do monitoramento, para o SGBD do sistema em avaliação;

- **Tabelas de Valores de Dimensões** são as tabelas em que ficam armazenados os valores referentes ao monitoramento das transações para futura emissão dos relatórios;
- **Tabelas de Valores Criadas** é o retorno do SGBD do SI em avaliação, informando que as tabelas foram adicionadas à base desse SI;
- **Enviar Script para a Criação das Triggers de Monitoramento** é o envio de sinais em uma linguagem de programação SQL – PL/SQL – ao SGBD do sistema avaliado, para a criação das *triggers* de monitoramento na base de dados desse SI;
- **Triggers de Monitoramento** são as *triggers* responsáveis por monitorar as transações ocorridas com os dados em avaliação;
- **Triggers Criadas** é a mensagem de retorno do SGBD informando que as *triggers* foram adicionadas à base de dados do SI avaliado;
- **Monitorar as Transações nos Dados e Gravar Resultados** é a atividade de monitoramento das transações DML – Data Manipulation Language, mais precisamente *Insert*, *Update* e *Delete* ocorrida com os dados em avaliação e gravação dos resultados nas *tabelas de valores de dimensões*;
- **Requisitar Relatório** é a interação entre o avaliador e a ferramenta FEMEQ, solicitando a impressão de relatórios da qualidade;
- **Analisar Resultados** é a atividade em que o avaliador irá analisar os dados contidos nos relatórios acerca da qualidade dos dados monitoradas no SI para a sua certificação;
- **Propor Melhorias** é a atividade em que o avaliador irá fazer um diagnóstico acerca dos processos que estão gerando dados sem qualidade e propor alterações nos processos, a fim de melhorar a qualidade dos dados.

### 5.2.1 Adição de Entidades de Dimensão ao Modelo Lógico do Sistema em Avaliação

Para que seja possível a medição da qualidade de dados no sistema de informação em análise, com base no processo proposto, é necessário que sejam armazenados dados acerca da sua qualidade para a mensuração e análise. A FEMEQ baseia-se no conceito introduzido por Storey (1998), o qual determina que a ferramenta deve estar incorporada ao sistema em

avaliação, funcionando como uma extensão da base de dados do sistema. Isto permite que os dados sobre a qualidade sejam armazenados na própria base de dados do cliente, além de permitir o seu monitoramento por meio de *triggers* de banco de dados. A FEMEQ introduz na base de dados do SI em avaliação um conjunto de entidades que se relacionam com as entidades em que se encontram os dados que devem ser monitorados. Essas entidades são tabelas que guardam dados acerca das dimensões de qualidade e do resultado da avaliação dos dados monitorados para a geração de relatórios.

Para a ilustração dessa proposta, é utilizado um modelo lógico, apresentado na Figura 9, que demonstra a adição das entidades de dimensão na base de dados do sistema em avaliação. Nessa figura, a tabela realçada em amarelo faz parte da base de dados original do sistema em avaliação. As tabelas realçadas em cinza fazem parte das entidades adicionadas ao modelo original da base de dados do sistema avaliado.

Com o objetivo de evitar possíveis redundâncias com entidades já existentes no sistema em avaliação, todas as tabelas criadas pela FEMEQ começam com o prefixo “TES\$”, como pode ser visto na Figura 9. Essas entidades são descritas a seguir e o seu dicionário de dados se encontra no Apêndice A.

- Entidade TES\$DIMENSOES – armazena as dimensões que serão avaliadas na base de dados;
- Entidade TES\$AGREGACAO – responsável por estabelecer a relação entre a tabela de dimensões TES\$DIMENSOES e a tabela avaliada do SI. Identifica qual a tabela e atributo do SI que cada dimensão definida na atividade *Identificar Dimensões* do processo a ferramenta irá executar o monitoramento no sistema. Nela também é armazenada a data de início de avaliação da dimensão por atributo;
- Entidade TES\$VALORES\_<TABELA\_AVALIADA> – possui a função de armazenar os valores referentes ao resultado da avaliação dos dados contidos na tabela avaliada. <TABELA\_AVALIADA> corresponde ao nome da tabela que contém os dados em avaliação na base de dados avaliada. Na Figura 9, essa tabela é representada pela entidade “TES\$VALORES\_FUNCIONARIOS”. Neste caso, o nome da tabela é formado pela composição do prefixo “TES\$VALORES\_” com o nome da entidade original do SI em avaliação, “FUNCIONARIOS”.

É importante salientar que a adição dessas novas entidades na base de dados do sistema em avaliação não compromete os dados armazenados nessa base, nem no funcionamento do sistema em avaliação, já que não são alteradas as tabelas do modelo original.

Figura 9 - Modelo Lógico do Sistema em avaliação com as alterações realizadas pela FEMEQ



Fonte: O Autor

Para o exemplo ilustrado na Figura 9, foi gerado pela ferramenta, o *script* que se encontra no Apêndice D, Seção I, utilizando a DDL do padrão SQL.

### 5.2.2 Monitoramento dos Dados

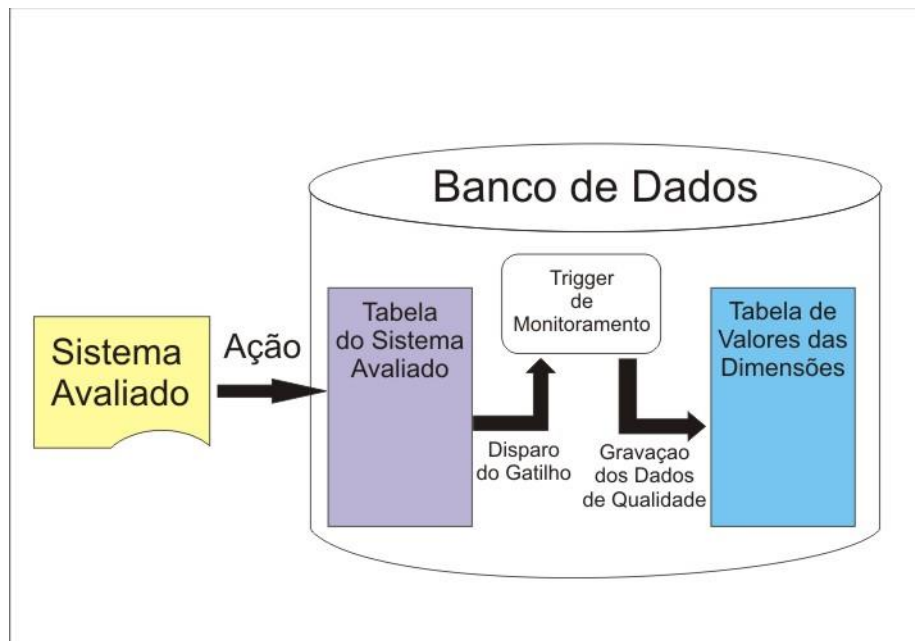
Pressman (2006) afirma que um software deve atender às necessidades e valores do negócio. Como cada SI possui suas próprias regras de negócio, a FEMEQ precisa ser flexível o suficiente para permitir a sua aplicação em diferentes SIs. Para satisfazer esse requisito, a FEMEQ faz uso de *triggers* que atuam como um “serviço” de monitoramento de transações,



já que a implementação de uma *trigger* pode ser realizada de acordo com as regras definidas para a avaliação da dimensão. Portanto, o avaliador deve definir essas *triggers*, baseado nas regras de negócio do SI em avaliação levantadas durante a elaboração da metodologia para a identificação das dimensões. A FEMEQ irá então adicionar as *triggers* definidas à base de dados do sistema avaliado. É importante ressaltar que a ferramenta já apresenta alguns modelos para a criação dessas *triggers*, além da possibilidade do usuário adicionar novos modelos à ferramenta, porém esse trabalho deve ser feito de maneira individualizada para cada dimensão avaliada, por profissional que tenha conhecimento técnico de programação PL/SQL, a fim de evitar possíveis erros nos resultados.

A Figura 10 mostra a atuação das *triggers* de monitoramento na base de dados do sistema em avaliação.

**Figura 10 - Atuação da Trigger na Base de Dados do Sistema Avaliado**



Fonte: O Autor

Ao ser adicionada à base de dados do sistema avaliado, a *trigger* passa a monitorar todas as operações realizadas com os dados definidos na implementação dessa *trigger*. Sempre que ocorre uma ação (inserção, atualização ou exclusão) na tabela em que se encontra o dado monitorado, um gatilho é disparado e o código PL/SQL, definido em formato de *trigger*, é executado realizando a avaliação do conteúdo desse dado e o resultado dessa avaliação é armazenado na tabela de valores das dimensões.

### 5.2.3 Geração dos Relatórios

Na impressão dos relatórios de avaliação da qualidade de dados, a FEMEQ utiliza medidas da estatística descritiva para permitir a análise dos resultados do monitoramento da qualidade do dado de acordo com a dimensão avaliada. São apresentados nos relatórios a quantidade de registros monitorada, a soma dos valores inseridos durante o monitoramento, média dos valores armazenados, a soma dos quadrados dos desvios, a variância, o desvio padrão e o coeficiente de variação, resultantes do monitoramento, cujos cálculos estão descritos a seguir e baseados na seguinte notação:

- VAL\_VALOR: é o campo calculado da tabela de valores referente à dimensão de qualidade em avaliação;
- “n”: corresponde ao número de registros testados;
- <TABELA\_VALORES>: é a tabela cujos valores resultantes da avaliação dos dados são armazenados, conforme discutido na Seção 5.2.1;
- VAL\_POS\_CODIGO: é o atributo chave da tabela TESS\$AGREGACAO, descrita na Seção 5.2.1;
- <VALOR\_POS\_CODIGO>: refere-se ao valor do atributo chave da dimensão testada na tabela TESS\$AGREGACAO.

I. A *quantidade de registros monitorada* é calculada pela FEMEQ utilizando a função *count* da linguagem SQL;

II. A *Soma* é a somatória dos valores resultantes do monitoramento. Esses valores são referentes aos indicadores definidos na métrica durante a definição da metodologia.

$$\mathbf{Soma} = \sum VAL\_VALOR$$

III. A *Média* é a variável da estatística que irá calcular a divisão da soma dos valores resultantes do monitoramento pela quantidade de registros monitorados. “Sob uma visão geométrica a média de uma distribuição é o centro de gravidade, representa o ponto de equilíbrio de um conjunto de dados.” (GUEDES, 2014).

$$\mathbf{Media} = \frac{Soma}{n}$$

- IV.** A *Soma dos Quadrados dos Desvios* é a variável da estatística descritiva calculada pelo quadrado da soma de todos os valores avaliados, reduzidos da média. Servindo como base para o cálculo da variância, essa variável se mostra interessante porque, segundo Guedes, (2014), se o valor mais provável é a média aritmética das medidas, a soma dos quadrados dos desvios é um mínimo.

$$\text{Soma dos Quadrados dos Desvios} = \sum (VAL\_VALOR - MEDIA)^2$$

No qual, VAL\_VALOR é o campo calculado da tabela de valores referente à dimensão de qualidade em teste e MÉDIA, a soma dos valores VAL\_VALOR sobre a quantidade de elementos.

- V.** A *Variância* é a soma dos quadrados dos desvios dividida pelo número de elementos avaliados. Seu principal objetivo é permitir a análise do grau de variabilidade de determinadas amostragens e, através dela pode-se perceber desempenhos iguais, muito próximos ou muito distantes.

$$\text{Variância} = \frac{\text{Soma\_do\_quadrado\_dos\_desvios}}{n}$$

No qual, “n” representa o número de elementos avaliados.

- VI.** O *Desvio Padrão* é uma medida que mede a dispersão dos valores em torno da média. Ele é usado para demonstrar o afastamento do conjunto de dados analisados em relação à média.

$$\text{Desvio Padrão} = \sqrt{\text{Variância}}$$

- VII.** O *Coefficiente de Variação* é utilizado para estimar a precisão de experimentos e representa o desvio-padrão expresso como porcentagem da média. “A partir do coeficiente de variação pode-se avaliar a homogeneidade do conjunto de dados e, conseqüentemente, se a média é uma boa medida para representar estes dados” (GUEDES, 2014).

$$\text{Coeficiente de variação} = \frac{\text{desvio padrão}}{\text{média}} \times 100$$

Com essas variáveis estatísticas, é possível realizar uma análise comparativa dos valores gerados, identificar os problemas de qualidade de acordo com as métricas definidas. Para se realizar uma análise estatística pode ser usado o coeficiente de variação, pois ele leva em consideração a medida de dispersão absoluta (desvio padrão) e a média da série. Portanto a medida de dispersão relativa (coeficiente de variação) prevalece sobre a medida de dispersão absoluta (desvio padrão).

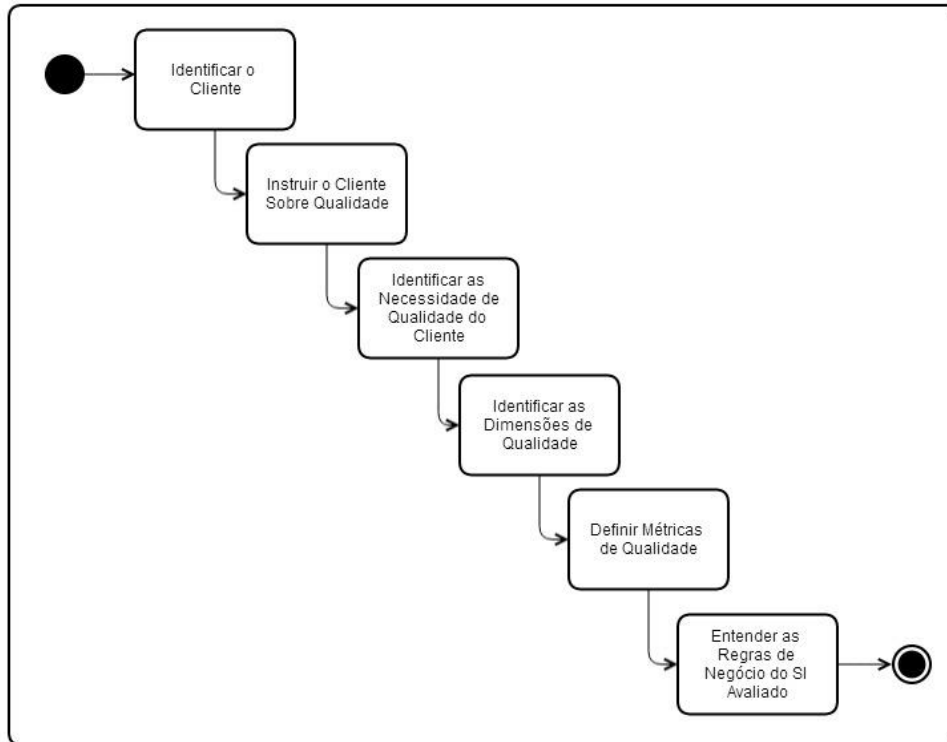
#### **5.2.4 Uma Proposta para a Definição das Dimensões de Qualidade sob a Perspectiva do Usuário**

No Capítulo 4, são referenciadas algumas metodologias utilizadas por pesquisadores para determinar quais são as necessidades de qualidade da informação para SI. Essa seção apresenta um modelo de processo para auxiliar o avaliador na definição das dimensões de qualidade na percepção do cliente, usuário do SI.

É importante ressaltar que os usuários de SI sabem que necessitam de qualidade, porém, muitos deles não sabem como determiná-la. O'Brien (2004) define os usuários de um SI como sendo as pessoas que utilizam um sistema de informação ou as informações que ele produz. Portanto, é necessário conhecer o que esse usuário espera do SI em termos de qualidade da informação para servir de base para a determinação de quais dimensões de qualidade devem ser aplicadas aos dados para a sua certificação. É visível então, que sob essa perspectiva, a participação do usuário mostra-se como uma importante contribuição para melhoria da qualidade dos sistemas de informação.

A Figura 11 ilustra um modelo de processo para a determinação das dimensões de qualidade, cujas atividades são explicadas a seguir.

**Figura 11 - Determinação das Dimensões de Qualidade de Dados**



Fonte: O Autor

- **Identificar o Cliente:** antes de inicializar definir as dimensões de qualidade, é necessário identificar quem será o usuário do SI, aqui chamado de *cliente*, que fará uso das informações geradas por ele;
- **Instruir o Cliente Sobre Qualidade:** como o cliente é o mais interessado na melhoria da qualidade nas informações que utiliza, sendo um participante ativo do processo de avaliação, é importante que o avaliador o oriente sobre como ocorrerá todo o processo de avaliação, explicando o que é QI e como ela será avaliada, além de como os resultados serão apresentados. Assim, não será criada nesse usuário falsas expectativas acerca do processo de melhoria da QI.
- **Identificar as Necessidades de Qualidade do Cliente:** nesta fase, com o cliente devidamente instruído, são levantadas as necessidades de qualidade sob a sua perspectiva. Esse procedimento é importante, haja vista que as necessidades de qualidade podem variar de um cliente para outro, mesmo estando utilizando o mesmo SI. A identificação das necessidades dos clientes pode ser realizada através da aplicação ferramentas que permitam identificar as percepções da qualidade de dados dos profissionais que utilizam as informações geradas pelo SI, a exemplo de questionários e entrevistas;

- **Identificar as Dimensões de Qualidade:** conhecidas as necessidades do cliente, é imperativo identificar quais as dimensões de qualidades serão aplicadas em quais dados no SI a ser avaliado;
- **Definir Métricas de Qualidade** para facilitar a avaliação da qualidade, é importante definir métricas para cada dimensão avaliada, criando uma escala de valores possíveis para os resultados da avaliação dos dados em cada dimensão, o que permite a quantificação dos resultados dos testes realizados nos dados;
- **Entender as Regras de Negócio do SI Avaliado:** é necessário conhecer as regras de negócio do sistema avaliado, identificando os processos que manipulam e geram os dados que estão sendo avaliados. Esse processo também auxiliará na identificação fase de proposição de melhorias apresentadas na TDQM.

Durante o processo de avaliação da qualidade, esse modelo deve ser instanciado para cada cliente afim de que a sejam identificadas as dimensões de qualidade que devem ser aplicadas ao conjunto de dados avaliados e definidas as regras de negócio para a criação das *triggers* de monitoramento.

### 5.3 CONSIDERAÇÕES FINAIS

Neste capítulo foram abordados os principais aspectos sobre a implementação e utilização da FEMEQ. Foram, também, propostos um modelo para a medição da qualidade de dados utilizando a ferramenta FEMEQ, bem como um modelo de processo para a definição dos dados, dimensões e métricas de avaliação da qualidade sob a perspectiva do usuário. Esses modelos devem ser instanciados para cada sistema em avaliação para facilitar a utilização da ferramenta. O modelo de definição das dimensões sob a perspectiva do usuário permite ao avaliador entender as regras de negócio do sistema em avaliação, bem como identificar os dados a serem avaliados, além das dimensões e métricas para avaliação, de acordo com a necessidade do usuário do SI avaliado, principal interessado nos resultados da avaliação.

O conceito de adição de entidades de dimensões no modelo de dados do sistema a ser monitorado para a avaliação da qualidade dos dados já foi utilizado em outros trabalhos, a

exemplo de Favaretto (2007) e Amaral (2003), demonstrando a sua eficácia. O processo de monitoramento por transações utilizando *trigger* garante uma avaliação contínua dos dados, fazendo uso dos recursos do próprio SGBD do sistema avaliado, diferentemente do monitoramento utilizando serviços externos. O grande desafio do avaliador está na criação de uma metodologia criteriosa para a definição das *triggers*, sem que sejam gerados erros de execução e, conseqüentemente, discrepância nos resultados a serem analisados.

O Capítulo 6 apresenta a aplicação dos modelos aqui propostos com a utilização da FEMEQ, através de estudos de caso.

## CAPITULO 6 – Estudos de Caso Utilizando a FEMEQ

### 6.1 INTRODUÇÃO

Com o intuito de avaliar o processo e ferramenta propostos no Capítulo 5, neste capítulo serão apresentados três estudos de caso, cada um com suas especificidades, como definidos a seguir, para avaliar a qualidade dos dados de diferentes SI com o auxílio do processo e ferramenta propostos nesta dissertação.

Entende-se que a utilização de estudos de caso é útil para investigar novos conceitos, bem como verificar como são aplicados e utilizados na prática elementos de uma teoria (YIN, 2009 *apud* BRANSKI *et al*, 2010). Sendo assim, estudos de caso são aplicados neste capítulo para demonstrar as potencialidades da ferramenta FEMEQ, bem como comprovar a sua eficiência, além de validar o processo de avaliação da qualidade de dados baseado na ferramenta proposta.

É importante ressaltar que neste capítulo, apenas serão executadas as fases de *definição da qualidade, coleta de dados e análise dos resultados* definidas pela TDQM, de maneira que se possa avaliar o funcionamento da ferramenta com o propósito de atender o objetivo definido para o desenvolvimento desta dissertação, i.e. a análise da qualidade dos dados de acordo com a percepção dos usuários dos sistemas em avaliação.

### 6.2 APRESENTAÇÃO DOS CENÁRIOS

As avaliações foram feitas a partir de três diferentes clientes que atuam em áreas distintas e possuem software para o gerenciamento de suas atividades. A metodologia de utilização da FEMEQ foi aplicada nas bases de dados de cada um desses clientes. Todos três sistemas utilizam o Sistema Gerenciador de Banco de Dados Firebird. Por solicitação dos clientes, tanto os seus dados, como os de suas respectivas empresas não devem ser publicados, sendo assim, neste trabalho, serão denominados de *Cliente A*, *Cliente B* e *Cliente C*.

A seguir são descritos os três ambientes de avaliação:

- (i) O Cliente A é uma academia de musculação com aproximadamente 2.100 (dois mil e cem clientes), com uma rotatividade de aproximadamente 600 (seiscentos)



clientes/mês. Foi avaliado um software que realiza a gestão financeira e acadêmica dos alunos. Nesse cenário existe um processo automatizado de carga de dados para o registro da frequência dos alunos através de uma catraca eletrônica integrada ao sistema. Foi entrevistado o usuário das informações gerenciais geradas pelo sistema;

- (ii) O Cliente B é um hotel contendo 35 apartamentos e com uma rotatividade de aproximadamente 255 (duzentos e cinquenta e cinco) clientes/mês. Foi avaliado um software ERP que integra atividades que vão desde a hospedagem até o CRM – Customer Relationship Manager, incluindo o controle financeiro e de telefonia. Nesse cenário existem dois processos críticos que são o checkin e checkout de hóspedes, cujas as informações devem ser precisas para viabilizar a utilização das informações de CRM. Além desse, existe também um processo onde os dados de telefonia são carregados automaticamente através de uma central telefônica integrada ao sistema. Nesse cliente, o provedor de informações foi o gerente administrativo, um usuário que possui um amplo conhecimento acerca desse sistema de informação:
- (iii) O Cliente C é um clube com mais de 4000 (quatro mil) associados, com uma rotatividade de 1300 (um mil e trezentos) associados no período de alta temporada. Nesse cenário existe um processo crítico que é a geração automática, em lote, de boletos de cobrança. A avaliação foi realizada no sistema que serve para auxiliar o gerenciamento de mensalidades e acesso de associados ao clube. Para a identificação dos dados para avaliação da qualidade do SI foi entrevistado o seu presidente, que possui amplo conhecimento acerca do ambiente informacional da empresa.

### 6.3 A METODOLOGIA DE AVALIAÇÃO

Como especificado no processo de avaliação da qualidade de SI proposto no Capítulo 5, Seção 5.2, uma metodologia deve ser definida para a identificação da definição das dimensões de qualidade sob a perspectiva do cliente, o que servirá como base para a definição de quais dimensões de qualidade deverão ser aplicadas em cada atributo e tabela do SI a ser avaliado, além de permitir a criação das *triggers* de monitoramento automatizado quando for conveniente.

Sendo assim, para a identificação das dimensões, foi utilizado o modelo de metodologia proposto na Seção 5.2.4, cujas atividades são descritas a seguir:

## **I. Identificar o Cliente**

Foi realizada uma entrevista com os principais usuários do SI de cada cliente, na qual foi identificado o principal utilizador das informações geradas pelo sistema.

## **II. Instruir o Cliente Sobre Qualidade**

Depois de identificado o principal usuário do SI em avaliação, foi discutido e definido com os clientes o que é qualidade de software, como ela deve ser medida e quais os resultados esperados na avaliação. Foi explicado também que uma baixa qualidade de dados pode advir não somente de falha nos sistemas em avaliação, mas também dos processos de carga e manipulação dos dados desses sistemas. Ademais, foi entregue aos clientes, juntamente com o questionário, um documento de apresentação explicando o objetivo da pesquisa e as instruções de preenchimento do questionário, como pode ser vista no Apêndice C.

## **III. Identificar as Necessidades de Qualidade do Cliente**

Foi então aplicado, para os três clientes, um questionário solicitando que cada um apontasse ao menos três aspectos considerados críticos em seus SI e que eles desejassem que tivessem a qualidade avaliada. A resposta deveria ser descritiva e, para cada item listado, seria informado o grau de relevância desses aspectos para o seu sistema, segundo a sua visão. Poderia ser aceito um dos três valores para o grau de relevância: Baixa, Média ou Alta. Isto serviu para identificar quais processos são mais importantes para determinar a qualidade do sistema. Esse formulário pode ser encontrado no Apêndice C. As respostas ao questionário foram transcritas fielmente como os usuários responderam e podem ser vistas na Tabela 5.

## **IV. Identificar as Dimensões da Qualidade**

Com base nas respostas obtidas, foram identificadas as dimensões de qualidade e os atributos cujas dimensões seriam aplicadas. Os resultados dessa análise podem ser vistos na Tabela 5.

Tabela 5 - Dimensões apuradas para o Cliente A, Cliente B e Cliente C

Cliente	Aspecto	Dimensão	Grau de Relevância
A	“Preciso saber se na hora que cancelo a venda de uma mensalidade, a validade da mensalidade do aluno volta à data anterior”	Credibilidade da função de estorno da mensalidade	Alta
	“Quero que os tipos dos cartões sejam informados quando for realizada uma venda por cartão”	Completeza na venda com cartão	Alta
	“As vendas só podem ser canceladas pelos administradores”	Segurança no cancelamento das vendas	Média
B	“Verificar se a telefonia está sendo totalmente lançada”	Credibilidade no lançamento da telefonia	Alta
	“Verificar se o horário de entrada do cliente coincide com a hora cadastrada”	Temporalidade na hora de entrada na hospedagem do cliente	Baixa
	“Verificar se quando um aluguel é cancelado, o sistema está armazenando o nome do funcionário que cancelou”	Completeza no cancelamento da reserva	Média
C	“As saídas dos sócios e dependentes estão sendo registradas?”	Completeza no registro de frequência dos associados	Baixa
	“Está sendo registrado o código do funcionário nas liberações de entrada dos associados com ocorrência?”	Credibilidade no registro de entrada	Média
	“As mensalidades estão sendo lançadas no período certo?”	Temporalidade no lançamento da mensalidade	Alta
	“Quantos boletos estou tendo que emitir para cobrar uma mensalidade?”	Quantidade apropriada dos dados na função de emissão dos boletos.	Media

Fonte: O autor

## V. Definir Métricas de Qualidade

Depois de identificadas as dimensões, foram determinados os valores possíveis para cada uma delas, a fim de quantificá-las, definindo métricas para a sua medição. Wang et al (1996), afirmam que critérios subjetivos devem ser suprimidos na medição da qualidade da informação e, baseado nessa afirmação, foram definidos apenas dois valores para cada dimensão, como mostrado na Tabela 6 e definidos o valor 0 para os casos negativos e 1 para os casos positivos, como métrica para a quantificação da qualidade.

**Tabela 6 - Relação entre dimensões, valores e métricas utilizados para os testes da QI nos clientes**

<b>Dimensão</b>	<b>Valores</b>	<b>Métricas</b>
Credibilidade	Com Credibilidade	1
	Sem Credibilidade	0
Completeza	Completo	1
	Não Completo	0
Segurança	Seguro	1
	Não Seguro	0
Temporalidade	No tempo	1
	Fora do Tempo	0
Quantidade apropriada dos dados	Apropriada	1
	Não Apropriada	0

Fonte: O autor

## **VI. Entender as Regras de Negócio do SI Avaliado**

Essa é a fase mais trabalhosa do modelo proposto. Para a definição dos procedimentos armazenados que serão disparados pelas *triggers*, é necessário o conhecimento dos processos que realizam a carga dos dados na base de dados dos sistemas em avaliação. Então, com as dimensões definidas, foi realizado um estudo sobre o funcionamento de cada sistema. Para isso, utilizando técnica de observação, foi desenvolvida uma sinopse do funcionamento de cada processo utilizado para o carregamento dos dados nos sistemas e identificado nos principais usuários envolvidos nesses processos, as diversas maneiras de realizar as mesmas tarefas. Em seguida, foi utilizada como técnica de engenharia reversa, a extração dos metadados de cada banco de dados em avaliação, o que possibilitou a criação de um modelo entidade-relacionamento para cada SI, permitindo um estudo das entidades e procedimentos armazenados de banco de dados que abrigam e/ou manipulam os dados em avaliação de cada sistema.

Depois de aplicada a metodologia para identificação das dimensões de qualidade a serem avaliadas, as métricas para a medição da qualidade foram inseridas na ferramenta. Então, com a adição das entidades referentes às dimensões no modelo de dados dos clientes, ver Seção 6.4, foi possível armazenar os resultados referentes às medições da QI nos sistemas em avaliação, servindo como base para a geração dos relatórios que irão demonstrar os níveis de qualidade nos dados contidos nos respectivos repositórios de dados avaliados.

## 6.4 ADIÇÃO DAS TABELAS DE DIMENSÕES NO MODELO LÓGICO DE DADOS DOS CLIENTES

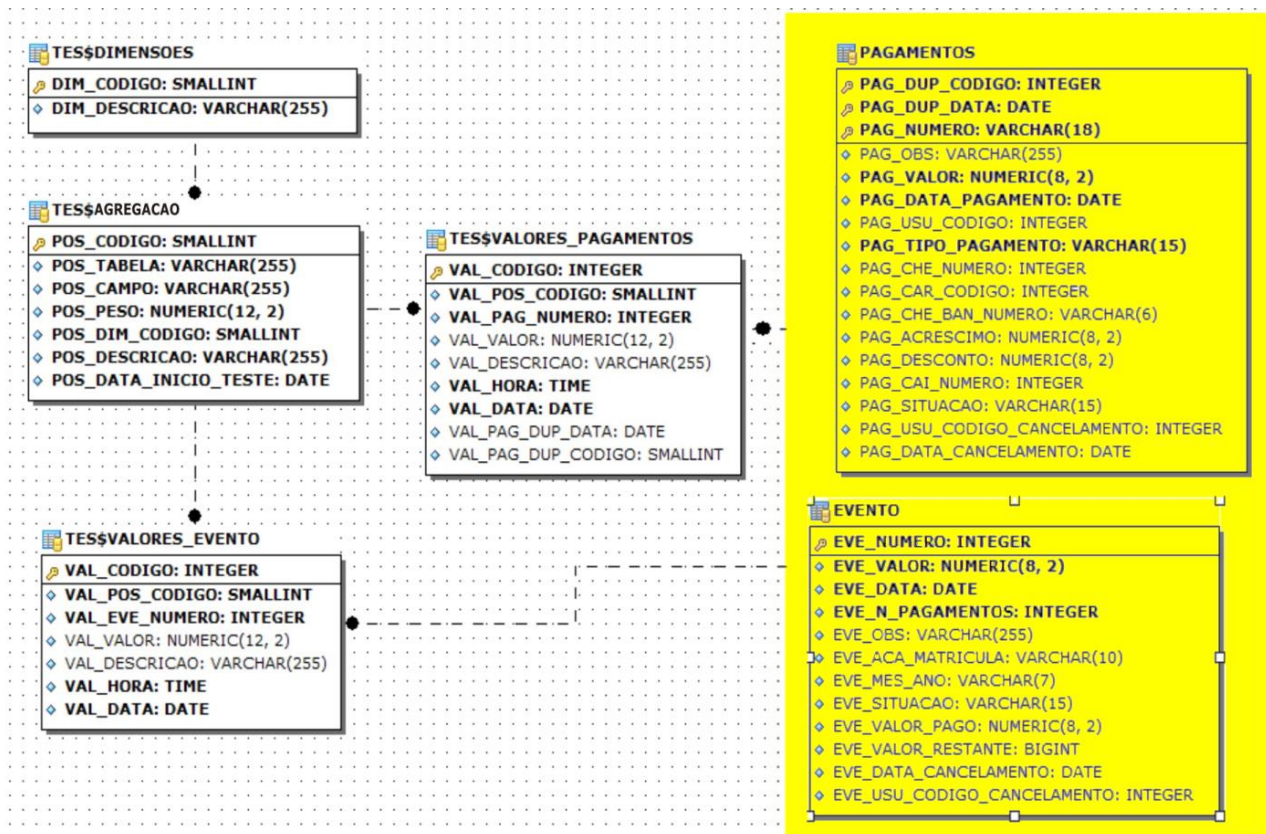
Como discutido no Capítulo 5, Seção 5.2.1, ao inserir as dimensões na FEMEQ são adicionadas ao modelo de dados do sistema em avaliação as tabelas para o armazenamento dos valores relacionados à qualidade de dados. A seguir são demonstradas as alterações realizadas no modelo lógico de cada cliente.

### 6.4.1 Cliente A

A Figura 12 apresenta o modelo lógico do Cliente A com as respectivas entidades do modelo original e as adicionadas pela FEMEQ. Realçada em amarelo, encontram-se as tabelas do modelo original utilizado pelo SI, as demais tabelas são as da FEMEQ utilizadas para realizar o monitoramento das transações.

Para o monitoramento das dimensões *Credibilidade da função de estorno da mensalidade* e *Completeza na venda com cartão*, vide Tabela 5, foram adicionadas pela ferramenta, além das tabelas padrão TESS\$DIMENSOES e TESS\$AGREGACAO, a tabela TESS\$VALORES\_PAGAMENTOS que será o repositório dos valores resultantes do monitoramento, relativos à medição da qualidade, de acordo com as métricas definidas na metodologia e aplicadas pelas regras de monitoramento definidas nas *triggers* dessas duas dimensões. A FEMEQ adiciona, para cada entidade em avaliação, uma tabela de valores, ao invés de adicionar uma tabela para cada métrica, com o objetivo de reduzir a quantidade de entidades acrescentadas à base de dados do sistema em avaliação, além de evitar a criação de entidades redundantes e desnecessárias.

Figura 12 - Modelo Lógico do Cliente A



Fonte: O Autor

Para a dimensão *Segurança no cancelamento das vendas*, foi adicionada a tabela TESSVALORES\_EVENTO que armazenará os valores relativos à qualidade dos dados para esta dimensão. Para cada dimensão avaliada, foi gerada uma *trigger* para realizar o monitoramento das transações do SI. O *script* de criação dessas *triggers* pode ser visto no Apêndice B, Seção I. Os *scripts* de geração das tabelas das dimensões supracitadas são mostrados no Apêndice D, Seção II.

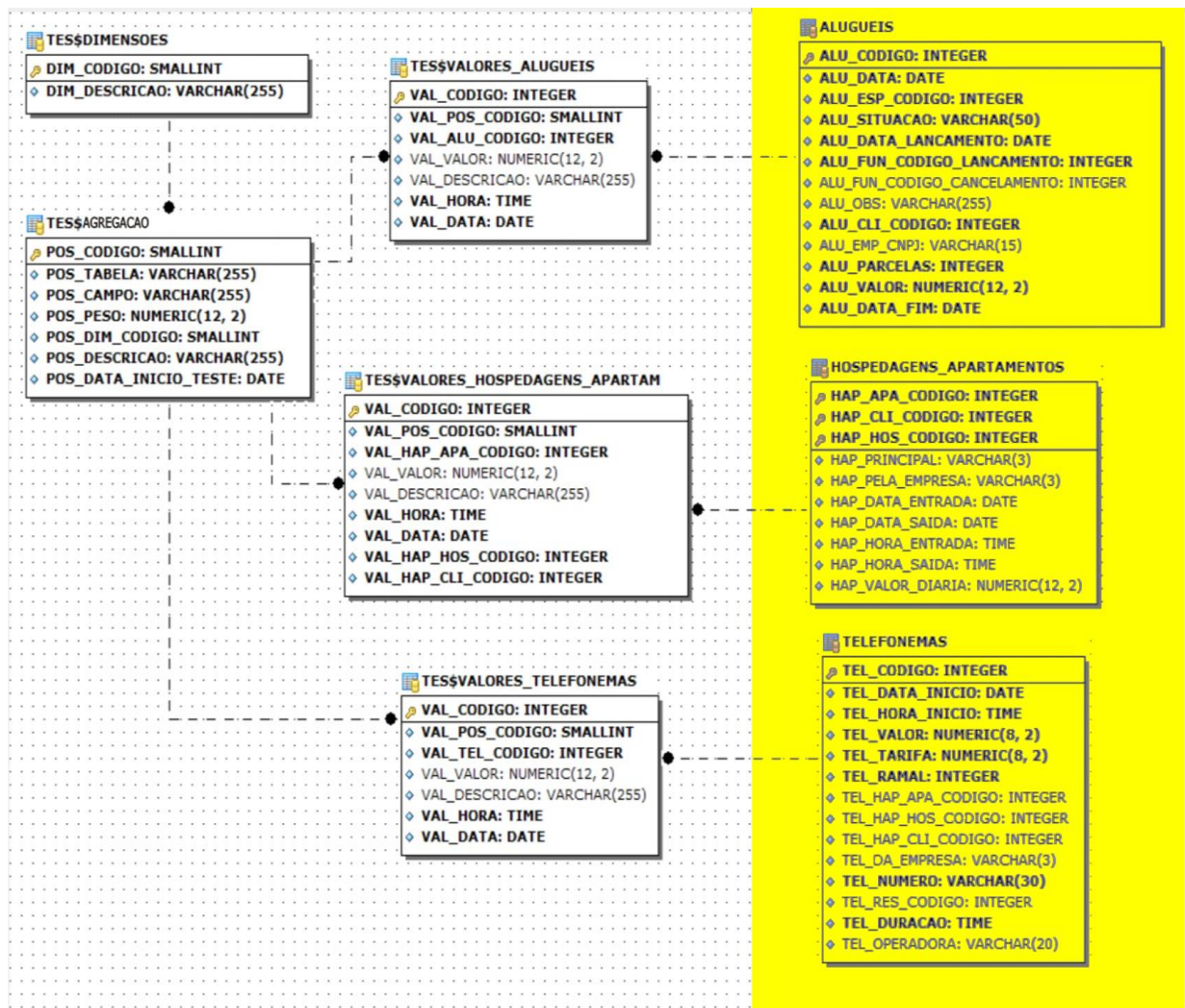
## 6.4.2 Cliente B

As alterações realizadas pela FEMEQ no modelo lógico do Cliente B são demonstradas na Figura 13. Nela é possível observar que foram adicionadas três tabelas de valores, além das tabelas de padrão de dimensões. Assim como no modelo do Cliente A, as tabelas originais do modelo estão realçadas em amarelo, sendo apresentadas apenas as que possuem relevância para a avaliação das dimensões selecionadas.

Observando o modelo é possível verificar que foram adicionadas as tabelas TESSVALORES\_ALUGUEIS para a dimensão *Completeza no cancelamento do aluguel*, TESSVALORES\_HOSPEDAGENS\_APARTAM para a dimensão *Temporalidade na hora de entrada na hospedagem do cliente* e TESSVALORES\_TELEFONEMAS para a dimensão *Credibilidade no lançamento da telefonia*. Todas as tabelas citadas têm a função de armazenar os valores resultantes do monitoramento das transações ocorridas com as tabelas do modelo original, as quais se relacionam. Assim como no Cliente A, foram criadas *triggers* de monitoramento para cada uma das dimensões avaliadas e o script de criação dessas *triggers* pode ser visto no Apêndice B, Seção II.

Os scripts de geração das tabelas TESSVALORES\_ALUGUEIS, TESSVALORES\_HOSPEDAGENS\_APARTAM e TESSVALORES\_TELEFONEMAS de dimensões supracitadas são demonstrados no Apêndice D, Seção III.

Figura 13 - Modelo Lógico do Cliente A

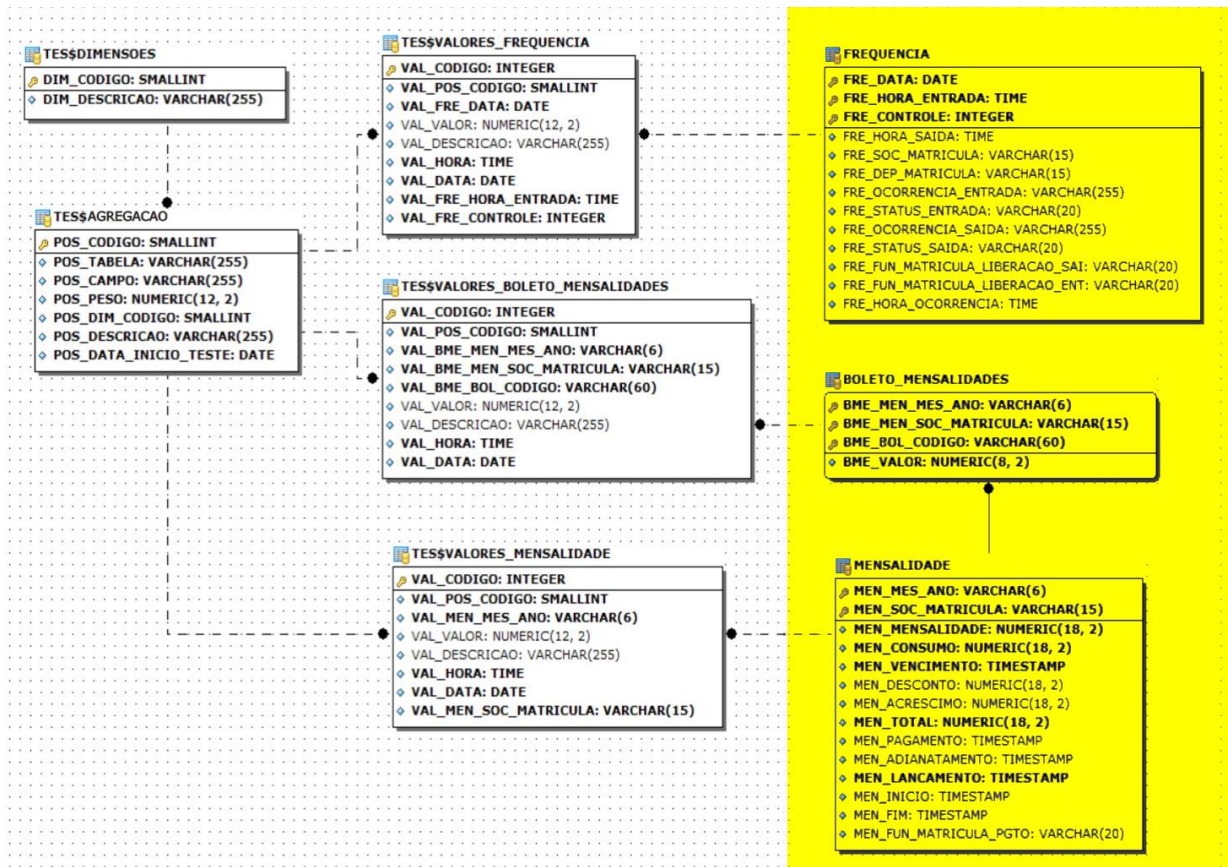


Fonte: O Autor

### 6.4.3 Cliente C

Na Figura 14 é possível visualizar as alterações realizadas no modelo lógico do Cliente C pela FEMEQ. É possível observar que, diferentemente do Cliente A, foram adicionadas três tabelas de valores, já que cada dimensão avaliada se refere a uma entidade diferente no modelo de dados do sistema, além das tabelas padrão de dimensões. As tabelas originais do modelo estão destacadas em amarelo e são apresentadas apenas as que possuem relevância para a avaliação das dimensões selecionadas.

Figura 14 - Modelo Lógico do Cliente C



Fonte: O Autor

Observa-se no modelo que a FEMEQ adicionou as tabelas TESSVALORES\_FREQUENCIA para as dimensões *Credibilidade no registro de entrada* e *Completeza no registro de frequência dos associados*, TESSVALORES\_BOLETO\_MENSALIDADES para a dimensão *Quantidade apropriada dos dados na função de lançamento dos boletos* e TESSVALORES\_MENSALIDADE para a dimensão *Temporalidade no lançamento da mensalidade*. Todas as tabelas citadas têm a



função de armazenar os valores resultantes do monitoramento das transações ocorridas com as tabelas do modelo original, as quais se relacionam. Assim como nos exemplos anteriores, foram criadas *triggers* de monitoramento para cada uma das dimensões avaliadas e o *script* de criação dessas *triggers* pode ser visto no Apêndice B, Seção III.

O *script* para criação das tabelas de valores TESSVALORES\_FREQUENCIA, TESSVALORES\_BOLETO\_MENSALIDADES, TESSVALORES\_MENSALIDADE é demonstrado no Apêndice D, Seção IV.

## 6.5 RESULTADOS OBTIDOS

De acordo o modelo de processo de avaliação da qualidade de dados descrito no capítulo anterior, cada cliente será analisado em todos os itens apontados como importantes para a qualidade de dados do seu sistema. Os testes foram realizados nos sistemas durante 30 (trinta) dias, período compreendido entre primeiro de janeiro de dois mil e quatorze e trinta e um de janeiro do mesmo ano. Os resultados foram sintetizados na Tabela 7 e serão detalhados nesta seção.

**Tabela 7 - Resultados obtidos pela FEMEQ**

Cliente	Dimensão	Registros	Soma	Média	Quadrado dos Desvios	Variância	Desvio Padrão	Coefficiente de Variação
A	Credibilidade no estorno da mensalidade	7	0.00	0.00	0.00	0.00	0	0
	Completeza na venda com cartão	134	126	0.94	7.52	0.06	2.74	25.19
	Segurança no cancelamento das vendas	14	14	1.00	0.00	0.00	0	0
B	Credibilidade no lançamento da Telefonia	262	138	0.52	65.32	0.25	8.08	96.02
	Temporalidade na hospedagem do cliente	240	237	0.98	2.98	0.012	1.73	11.36
	Completeza no cancelamento de aluguel	3	2.00	0.66	0.66	0.22	0.82	71.42
C	Completeza no Registro de frequência dos associados	1054	395	0.37	246.99	0.23	15.72	130.82
	Credibilidade no registro de entrada de sócios	178	178	1.00	0.00	0	0	0
	Temporalidade no lançamento da mensalidade	1485	1482	0.99	3.09	0.002	1.76	4.52
	Quantidade apropriada de dados em boleto	634	631	0.99	3.00	0.005	1.73	6.92

Fonte: O Autor – Extraído da Ferramenta FEMEQ

### 6.5.1 Cliente A

Analisando os resultados do Cliente A, que podem ser visualizados na Figura 15, para a dimensão de *Credibilidade da função de estorno da mensalidade* foram testados sete registros, que apesar de não apresentar uma quantidade significativa de dados, se justifica porque, segundo informações do cliente, não é uma prática tão comum estornar mensalidade na academia. Como resultado, a média, soma, variância, desvio padrão e coeficiente de variação retornaram o valor zero, o que demonstra, de acordo a métrica definida na Tabela 6 que todos os registros obtiveram o valor zero, demonstrando que em nenhum dos casos testados a credibilidade foi confirmada. Portanto, com base nesses resultados pode-se inferir que a qualidade dos dados geradas no processo de estorno da mensalidade pode estar comprometida. Ainda que não seja o objetivo deste trabalho diagnosticar a causa da falha durante o processo que resultou na baixa qualidade dos dados, mas a empresa desenvolvedora do sistema foi consultada e esta informou que o software não está pronto para atender a esse requisito, embora o manual do sistema entregue ao usuário informe o contrário.

**Figura 15 - Resultado dos testes de qualidade para a credibilidade no estorno de mensalidade**

CNPJ:	Fones:	/	Fax:			
Endereço:						
E-Mail:						
CREDIBILIDADE NO ESTORNO DE MENSALIDADE						02/02/2014:Data 20:21:Hora 1: Pág
Registros:	Soma:	Média:	Quadrado dos Desvios:	Variância:	Desvio Padrão:	Coeficiente de Variação:
7	0,00	0,00	0,0000	0,0000	0	0

Fonte: O Autor – Extraído da Ferramenta FEMEQ

Já para a dimensão *Completeza na venda com cartão*, pode ser observado na Figura 16, que foram testados 134 (cento e trinta e quatro) registros, com uma média de 0,94 para um desvio padrão de aproximadamente 2,74. Embora o coeficiente de variação possua um valor aproximado de 25%, o que sugere um resultado heterogêneo, a média apresentou um valor relativamente alto com uma pequena variância. Isso permite deduzir que a qualidade dos dados nesse item está aceitável, mas ainda não é a ideal, merecendo uma atenção especial na fase de melhoria da TDQM. Uma análise preliminar acerca das possíveis causas para o

problema da qualidade nesse conjunto de dados indicou que um usuário do sistema não informava o tipo do cartão durante o cadastro das vendas com cartão. Diante disso, o processo de carga de dados durante a venda deve ser revisto pela empresa desenvolvedora, inserindo como requisito a obrigatoriedade da informação do tipo cartão nas vendas por cartão.

**Figura 16 - Resultado dos testes de qualidade para a Completeza na venda com o Cartão**

CNPJ:		Fones:		/		Fax:	
Endereço:							
E-Mail:						02/02/2014:Data	
<b>COMPLETEZA NA VENDA COM CARTÃO</b>							20:19 :Hora
							1.:Pág
Registros:	Soma:	Média:	Quadrado dos Desvios:	Variância:	Desvio Padrão:	Coeficiente de Variação:	
134	126,00	0,94	7,5224	0,0561	2,74269940022599	25,1972750687809	

Fonte: O Autor – Extraído da Ferramenta FEMEQ

A terceira dimensão analisada foi *Segurança no cancelamento das vendas*, que, de acordo com a Figura 17, demonstrou um alto grau de qualidade, cuja soma calculada é exatamente igual ao de registros testados e o cálculo da média produziu o valor 1 (um), que de acordo com a Tabela 6, é considerado seguro.

**Figura 17 - Resultado dos testes de Segurança no cancelamento das vendas**

CNPJ:		Fones:		/		Fax:	
Endereço:							
E-Mail:						02/02/2014:Data	
<b>SEGURANÇA NO CANCELAMENTO DAS VENDAS</b>							20:17 :Hora
							1.:Pág
Registros:	Soma:	Média:	Quadrado dos Desvios:	Variância:	Desvio Padrão:	Coeficiente de Variação:	
14	14,00	1,00	0,0000	0,0000	0	0	

Fonte: O Autor – Extraído da Ferramenta FEMEQ

De acordo com os critérios definidos pelo usuário na Tabela 5, é correto afirmar que a baixa qualidade dos dados está impactando negativamente na qualidade do sistema, haja vista que as dimensões avaliadas, consideradas de alto e médio grau de relevância, apresentaram resultados negativos em termos de qualidade.

### 6.5.2 Cliente B

Para o cliente B, a avaliação da dimensão *Credibilidade no Lançamento da Telefonia*, apresentou resultados críticos e preocupantes, pois foram avaliados 262 registros e o

coeficiente de variação exibiu um valor relativamente alto, 96%, demonstrando uma amostragem altamente heterogênea, enquanto a média indicou que mais 50% dos dados foram classificados como sem qualidade, o que pode ser visto na Figura 18. Assim como no Cliente A, foi realizada uma investigação acerca do que poderia estar causando o problema e, depois de analisado o processo de carga dos dados da telefonia, descobriu-se que um problema em um equipamento de rede (hub) estava causando queda de comunicação entre o microcomputador responsável por receber os dados da central telefônica e o servidor de banco de dados, causando incompleta no processo e corrupção de dados. O diagnóstico dessa dimensão foi importante para este trabalho porque demonstra que os problemas referentes à qualidade de dados podem, também, advir de eventos externos ao sistema avaliado.

**Figura 18 - Resultado dos testes de credibilidade no lançamento da telefonia**

CNPJ:		Fones:		/		Fax:	
Endereço:							
E-Mail:							
CREDIBILIDADE NO LANÇAMENTO DA TELEFONIA							03/02/2014:Data
							11:59 :Hora
							1.:Pág
Registros:	Soma:	Média:	Quadrado dos Desvios:	Variância:	Desvio Padrão:	Coeficiente de Variação:	
262	138,00	0,52	65,3248	0,2493	8,08237588831403	96,0191364063071	

Fonte: O Autor – Extraído da Ferramenta FEMEQ

Quanto a avaliação da qualidade dos dados a partir da dimensão *Temporalidade na Hospedagem do Cliente* observa-se que o coeficiente de variação exibiu um valor baixo, em torno de 11%, o que indica uma amostra altamente homogênea e a média apresentou um resultado próximo a 1 (um), com um desvio padrão de 1,725 e variância de 0,0124, relativamente baixos, demonstrando um resultado satisfatório em termos de qualidade, como pode ser visto na Figura 19.

**Figura 19 - Resultado dos testes de Temporalidade na hospedagem do cliente**

CNPJ:		Fones:		/		Fax:	
Endereço:							
E-Mail:							
TEMPORALIDADE NA HOSPEDAGEM DO CLIENTE							03/02/2014:Data
							12:01 :Hora
							1.:Pág
Registros:	Soma:	Média:	Quadrado dos Desvios:	Variância:	Desvio Padrão:	Coeficiente de Variação:	
240	237,00	0,98	2,9760	0,0124	1,72510869222783	11,3627844139388	

Fonte: O Autor – Extraído da Ferramenta FEMEQ

A avaliação referente à dimensão *Completeza no Cancelamento do Aluguel*, Figura 20, não apresentou dados suficientes para uma análise mais precisa referente à esse processo, pois foram avaliados apenas 3(três) registros, porém, analisando os poucos dados coletados, é possível visualizar que existe um problema com esse processo, o que pode resultar em informações desconhecidas para a tomada de decisão pelos usuários do sistema. Apesar de esse processo ter gerado poucos registros, a média apresentou um resultado baixo 0,66, enquanto o coeficiente de variação apresentou um resultado bastante alto, cerca de 71%, o que demonstra um grau de heterogeneidade alto nessa amostragem. Para um resultado com maior relevância, é necessário que essa avaliação seja realizada em um período maior de tempo, o que permitirá a coleta de uma amostragem de dados maior.

**Figura 20 - Resultado dos testes para Completeza no cancelamento de aluguel.**

CNPJ:		Fones:		/	Fax:	
Endereço:						
E-Mail:						
COMPLETEZA NO CANCELAMENTO DE ALUGUEL						03/02/2014 :Data 11:58 :Hora 1.:Pág
Registros:	Soma:	Média:	Quadrado dos Desvios:	Variancia:	Desvio Padrão:	Coeficiente de Variação:
3	2,00	0,66	0,6668	0,2222	0,816578226503744	71,4213560569253

Fonte: O Autor – Extraído da Ferramenta FEMEQ

Deduz-se que o sistema de informações avaliado possui baixa qualidade dos dados, principalmente no processo de lançamento da telefonia, no qual, boa parte das ligações não está sendo cobrada. Apesar de demonstrar uma boa qualidade dos dados na dimensão de temporalidade na hospedagem do cliente, essa é considerada de baixa relevância. Já a completeza no cancelamento de aluguel precisa de uma amostragem maior de dados para um resultado mais preciso.

### 6.5.3 Cliente C

A avaliação realizada por meio da dimensão *Completeza no Registro de Frequência dos Associados*, para o cliente C, de acordo com a Figura 21, apresentou um coeficiente de variação muito alto, 130%, o que demonstra uma amostragem altamente heterogênea e uma média baixa, 0,37, o que comprova que a qualidade dos dados coletados é ruim, ou seja, mais da metade dos registros de frequência estão incompletos. Embora o cliente tenha classificado

essa dimensão como de baixa importância, os resultados indicam que ela implica na degradação da qualidade do sistema, haja vista que esses dados são utilizados na produção de diversos relatórios estatísticos para o auxílio à tomada de decisão. Foi realizado um levantamento do processo de carga dos dados de frequência do associado e foi observado que, apesar dessa carga ser realizada por um processamento automático com o auxílio de uma catraca+ eletrônica, a maioria dos usuários do clube utilizavam uma saída alternativa que não possui dispositivo de registro de entrada/saída, o que compromete os dados gerados nesse processo.

**Figura 21 - Resultado dos testes de Completeza no registro de frequência dos associados**

CNPJ:		Fones:		/		Fax:	
Endereço:							
E-Mail:							
<b>COMPLETEZA NO REGISTRO DE FREQUÊNCIA DOS ASSOCIADO</b>							04/02/2014:Data 00:55 :Hora 1.:Pág
Registros:	Soma:	Média:	Quadrado dos Desvios:	Variância:	Desvio Padrão:	Coeficiente de Variação:	
1054	395,00	0,37	246,9926	0,2343	15,7159982183761	130,823095246281	

Fonte: O Autor – Extraído da Ferramenta FEMEQ

A dimensão *Credibilidade no Registro de Entrada* apresentou na sua análise uma média alta e um coeficiente de variação igual a 0 (zero), como pode ser visto na Figura 22, o que demonstra um alto grau de qualidade nos dados gerados por esse processo.

**Figura 22 - Resultado dos testes de Credibilidade no Registro de Entrada de Sócios**

CNPJ:		Fones:		/		Fax:	
Endereço:							
E-Mail:							
<b>CREDIBILIDADE NO REGISTRO DE ENTRADA DE SOCIOS</b>							04/02/2014:Data 00:56 :Hora 1.:Pág
Registros:	Soma:	Média:	Quadrado dos Desvios:	Variância:	Desvio Padrão:	Coeficiente de Variação:	
178	178,00	1,00	0,0000	0,0000	0	0	

Fonte: O Autor – Extraído da Ferramenta FEMEQ

Observando a Figura 23, como resultado do monitoramento para a dimensão *Temporalidade no Lançamento da Mensalidade*, a ferramenta apresentou uma média alta, 0,99 e um coeficiente de variação baixo, aproximadamente 4,5%, e uma variância perto de 0 (zero), deduz-se, portanto, que a amostragem é homogênea e de alta qualidade.

**Figura 23 - Resultado dos testes de Temporalidade no lançamento da mensalidade**

CNPJ:	Fones:	/	Fax:						
Endereço:									
E-Mail:									
TEMPORALIDADE NO LANÇAMENTO DA MENSALIDADE									04/02/2014:Data 00:58 :Hora 1.:Pág
Registros:	Soma:	Média:	Quadrado dos Desvios:	Variância:	Desvio Padrão:	Coeficiente de Variação:			
1485	1482,00	0,99	3,0885	0,0020	1,75741287123999	4,51730904545412			

Fonte: O Autor – Extraído da Ferramenta FEMEQ

A Figura 24 mostra o resultado para a dimensão *Quantidade Adequada de Dados no Lançamento de Boletos* e a sua análise permite observar que a média apresentada está perto de 1 (um), com um coeficiente de variação de aproximadamente 7% e variância perto de 0 (zero), o que representa um conjunto de dados de alta qualidade.

**Figura 24 - Resultado dos testes para a Quantidade Adequada de Dados no lançamento de boletos**

CNPJ:	Fones:	/	Fax:						
Endereço:									
E-Mail:									
QUANTIDADE ADEQUADA DE DADOS EM BOLETO									04/02/2014:Data 00:57 :Hora 1.:Pág
Registros:	Soma:	Média:	Quadrado dos Desvios:	Variância:	Desvio Padrão:	Coeficiente de Variação:			
634	631,00	0,99	3,0034	0,0047	1,73303202509359	6,92490363676873			

Fonte: O Autor – Extraído da Ferramenta FEMEQ

Para o Cliente C, o sistema de informações, de acordo com as dimensões avaliadas, apresentou uma qualidade de dados satisfatória. Apenas a dimensão de *completeza no registro de frequência dos associados* demonstrou um resultado não desejado, mas como essa dimensão foi classificada pelo cliente como de baixa importância, a má qualidade não implica, necessariamente, numa degradação do sistema como um todo.

## 6.6 LIMITAÇÕES E RESTRIÇÕES DO ESTUDO DE CASO

A aplicação da ferramenta e do processo propostos neste trabalho, apresentado neste capítulo permitiu a medição da qualidade dos dados nos sistemas analisados, porém, algumas restrições podem ser observadas, como:

- O período de avaliação foi relativamente pequeno, não permitindo uma maior inferência acerca da qualidade em um número maior de transações;
- Não foram avaliados os dados em dimensões cuja medição necessite de intervenção humana, não podendo ser automatizada;
- Foi observado que o processo necessita de um intermediador que entenda os conceitos de banco de dados, representado no estudo pela figura do “avaliador”, para realizar a avaliação dos dados e apresentá-lo ao usuário final, no estudo, representado pelo “cliente”.

## 6.7 CONSIDERAÇÕES FINAIS

Neste capítulo foi apresentada uma aplicação da FEMEQ em três sistemas de informação e o processo de avaliação de qualidade de dados proposto nesta dissertação. Para aplicação dos testes, inicialmente foi definida uma metodologia de trabalho, quando foi aplicado um questionário para a identificação das dimensões de qualidade que deveriam ser aplicadas aos dados para satisfazer as questões de QI na percepção dos clientes. Em seguida, foram determinados os valores que cada dimensão poderia aceitar e definidas métricas para cada valor. Também foram especificadas as *triggers* para o monitoramento. Depois, as dimensões foram inseridas na FEMEQ, que adicionou tabelas de monitoramento no modelo de dados do cliente. Por fim, os resultados foram extraídos da ferramenta e analisados, apontando as dimensões que apresentaram qualidade e as que não apresentaram.

A aplicação do processo de avaliação proposto com o auxílio da FEMEQ permitiu a avaliação da qualidade de dados nos três cenários propostos e demonstrou que os problemas relacionados à qualidade de dados podem advir não só de problemas relacionados ao processamento realizado pelos sistemas de informação, mas também pela falta de orientação dos usuários, problemas em equipamentos, entre eles os de registro de dados e, até mesmo, de falhas na definição de requisito dos sistemas. A utilização desse processo, assim como da FEMEQ pode, então, auxiliar as empresas desenvolvedoras a proporcionar melhorias em seu processo de desenvolvimento de software, definição da infraestrutura necessária para o sistema e treinamento de usuários.

Apesar do período de avaliação ter sido relativamente pequeno, trinta dias, uma quantidade substancial de dados foi coletada na maioria dos sistemas, o que permitiu avaliar, na maioria dos casos, a qualidade dos dados gerados pelos sistemas transacionais e inferir a



origem e soluções para os problemas encontrados. Na Tabela 8 é possível visualizar os resultados obtidos com a avaliação dos sistemas de informações demonstrados neste capítulo.

**Tabela 8 - Resultados obtidos durante a avaliação**

	<b>Dimensão</b>	<b>Resultado</b>	<b>Problema Identificado</b>
<b>Cliente A</b>	Credibilidade da função de estorno da mensalidade	Sem Credibilidade	Implementação do Requisito
	Completeza na venda com cartão	Com Completeza	Utilização do Sistema / Implementação do Requisito
	Segurança no cancelamento das vendas	Com Segurança	Nenhum
<b>Cliente B</b>	Credibilidade no lançamento da telefonia	Sem Credibilidade	Infraestrutura
	Temporalidade na hora de entrada na hospedagem do cliente	No Tempo	Nenhum
	Completeza no cancelamento da reserva	Sem Completeza	Implementação do Requisito
<b>Cliente C</b>	Completeza no registro de frequência dos associados	Sem Completeza	Utilização do Sistema
	Credibilidade no registro de entrada	Com Credibilidade	Nenhum
	Temporalidade no lançamento da mensalidade	No Tempo	Nenhum
	Quantidade apropriada dos dados na função de emissão dos boletos.	Apropriada	Nenhum

Fonte: O Autor

## CAPITULO 7 - CONCLUSÕES E TRABALHOS FUTUROS

### 7.1 CONCLUSÕES

Neste trabalho foram apresentadas uma ferramenta para auxiliar a medição da qualidade de dados em ambientes transacionais e um processo de avaliação da qualidade dos dados baseado na percepção dos usuários. A ferramenta proposta deve ser genérica o suficiente para ser operacionalizada nos mais diversos ambientes de sistemas de informações transacionais.

Para consecução deste trabalho foi realizado uma revisão bibliográfica acerca do que é e como medir a qualidade de dados, na qual foi identificado que, para a medição da qualidade dos dados, se faz necessário o desenvolvimento de algumas atividades, como: a definição das dimensões de qualidade, os critérios de medição e análise dos resultados.

No Capítulo 5 foi apresentada a ferramenta e a proposta de um modelo de processo para a medição da qualidade dos dados baseada na perspectiva do usuário, fundamentada nas características encontradas nos trabalhos correlatos, discutidos no Capítulo 4, e consideradas relevantes para o desenvolvimento deste trabalho. Na Tabela 9 pode-se visualizar quais são essas características encontradas nos trabalhos correlatos e utilizadas pela FEMEQ e modelo de processo propostos.

Por seu caráter flexível, o modelo proposto pode utilizar os mais diversos algoritmos de avaliação, desde que possam ser adequados para a programação de procedimentos de banco de dados armazenados. Como pode ser visto na Tabela 9, é possível observar que todos os trabalhos apresentam a medição da qualidade dos dados, utilizam a definição das dimensões da qualidade e avaliam os dados com base na perspectiva do usuário final, apenas o trabalho de Pinho (2011) apresenta a avaliação dos dados baseados em não conformidades e somente o trabalho de Souza(2009) realiza a avaliação dos indicadores nos resultados, porém em nenhum deles é apresentada uma ferramenta que realize a avaliação dos dados de maneira automatizada e em tempo de transações, utilizando gatilhos (*triggers*) e procedimentos armazenados de banco de dados para isso.

A principal vantagem da utilização dos procedimentos armazenados é o aproveitamento dos recursos do próprio SGBD do sistema em avaliação, além da independência de um serviço externo para a avaliação dos dados e da flexibilização dos procedimentos de avaliação, como propõe o modelo apresentado.

Tabela 9 - Características para a avaliação de qualidade de dados

Trabalhos	Avaliação baseada em não conformidades	Medição da qualidade dos dados	Avaliação com base na perspectiva do usuário	Avaliação da apresentação dos indicadores nos resultados	Avaliação Automatizada, em tempo de transações, dos Dados	Definição de dimensões de qualidade
Amaral (2003)	Não	Sim	Sim	Não	Não	Sim
Calazans & Costa (2009)	Não	Não	Sim	Não	Não	Sim
Favaretto (2007)	Não	Sim	Sim	Não	Não	Sim
Kossowski & Favaretto (2010)	Não	Não	Sim	Não	Não	Sim
Kovac <i>et al</i> (1997)	Não	Sim	Sim	Não	Não	Sim
Lee <i>et al</i> (2001)	Não	Sim	Sim	Não	Não	Sim
Pinho(2001)	Sim	Sim	Sim	Não	Não	Sim
Silva e Loh (2011)	Não	Sim	Sim	Não	Não	Sim
Souza (2009)	Não	Sim	Sim	Sim	Não	Sim
FEMEQ e Modelo de Processo	Não	Sim	Sim	Não	Sim	Sim

Fonte: O Autor

No Capítulo 6, foram realizadas avaliações em três sistemas de informações diferentes para verificar a aplicabilidade e viabilidade técnica de uso do modelo de processo e da FEMEQ. Para a implantação do modelo de processo, foi realizada uma entrevista com os clientes, principais usuários de cada sistema testado, o que permitiu a identificação das dimensões críticas para cada SI, de acordo com as definições da TDQM, sendo definidas

métricas para cada uma delas e, finalmente, aplicadas na ferramenta propostas, a qual foi utilizada na avaliação dos sistemas de informação.

O processo apresentado no Capítulo 5 se mostrou eficiente quando aplicado nos estudos de caso discutidos no Capítulo 6, cuja análise dos dados foi executada sem falhas e foi possível inferir sobre os problemas na qualidade dos dados, identificando as suas possíveis causas. O modelo proposto no Capítulo 5 permitiu a definição das dimensões de qualidade sob a perspectiva do usuário durante as avaliações apresentadas no Capítulo 6. A avaliação da qualidade sob a perspectiva do usuário é importante porque ele é o principal interessado nas informações decorrentes do processamento dos dados realizados pelos sistemas, para a tomada de decisão.

Apesar de não estar no escopo deste trabalho, uma rápida investigação acerca das possíveis causas da baixa qualidade de dados observadas em alguns processos dos estudos de caso apontados no Capítulo 6 foi realizada e percebeu-se que alguns desses problemas tiveram origem em falhas na definição de requisitos dos sistemas, o que demonstra que a FEMEQ e o modelo de processo propostos podem ser importantes ferramentas para que os desenvolvedores de sistemas avaliem a qualidade dos seus processos de desenvolvimento de software e os otimizem.

O fato da ferramenta e do processo proposto no Capítulo 6 terem sido aplicados em três diferentes sistemas de informação demonstra a generalidade de aplicação. Porém, a sua comprovação deverá ser feita em uma quantidade maior de sistemas, o que deverá acontecer em trabalhos futuros.

### **7.1.1 Principais Contribuições**

Uma das contribuições deste trabalho perpassa pelo desenvolvimento de uma ferramenta, FEMEQ, que auxilie no monitoramento e avaliação da qualidade de dados em ambientes transacionais, baseada nos conceitos da TDQM, utilizando o conceito de dimensões de qualidade de dados. Outra importante contribuição foi o modelo de processo proposto, o qual permite aos avaliadores poderem mensurar a qualidade de dados com maior facilidade e confiabilidade.

Portanto, dentre as contribuições deste trabalho, podem ser citadas:

- Estudo e análise comparativa sobre as metodologias de definição de qualidade de dados;

- Avaliação da qualidade dos dados em tempo de execução das transações, com o uso de *triggers* como tecnologia para execução da avaliação;
- Proposta e implementação de uma ferramenta de medição de qualidade de dados baseada nos conceitos de dimensões de dados;
- Proposta de um modelo de processo para a avaliação da qualidade de dados sob a perspectiva do usuário;
- Proposta de um modelo de processo para a definição das dimensões de qualidade sob a perspectiva do usuário;
- Avaliação do processo de desenvolvimento, infraestrutura e utilização dos sistemas utilizados no estudo de caso.

### 7.1.2 Trabalhos Futuros

Dentre as sugestões para trabalhos futuros para o aperfeiçoamento da ferramenta e do processo propostos nesta dissertação, pode-se destacar:

- (i) Com a finalidade de ratificar a generalidade da ferramenta é importante realizar a avaliação da qualidade de dados em outros sistemas de informações, com diferentes particularidades e especificidades;
- (ii) A aplicação da ferramenta em base de dados maiores, com maior número de transações permitirá avaliar o desempenho da ferramenta em sistemas grandes;
- (iii) Também é importante avaliar os custos para a aplicação da ferramenta, já que o teste de qualidade consome tempo e trabalho, para verificar a viabilidade do seu emprego;
- (iv) O processo de avaliação proposto nesta dissertação, com a utilização da FEMEQ, realiza o monitoramento individualizado de todos os registros manipulados pelo sistema de informação avaliado. E, para cada registro avaliado, um novo registro é adicionado nas tabelas de valores para as dimensões em avaliação, o que pode aumentar consideravelmente o tamanho da base de dados do sistema. Sendo assim, uma linha de pesquisa que deve ser considerada, é a avaliação do custo em termos de espaço na base de dados que será consumido para realizar a avaliação dos dados nessa base;
- (v) Não foram discutidas neste trabalho, as situações em que as dimensões de qualidade não permitam a avaliação automatizada dos dados, apesar da ferramenta

ter sido projetada prevendo essa situação, isto deverá ser tratado em trabalhos futuros;

- (vi) Outra linha de pesquisa é a aplicação de outros processos de avaliação, diferentes do proposto nesta dissertação, para a avaliação da qualidade de dados em ambientes transacionais, fazendo uso da ferramenta FEMEQ;
- (vii) Não foi intenção deste trabalho apresentar planos de intervenção para os estudos de caso realizados a fim de se localizar os processos que estão causando problemas na qualidade dos dados nos sistemas avaliados, todavia, o modelo de processo pode ser estendido para realizar de forma sistemática e com uma metodologia definida a intervenção nos sistemas objetivando melhorias na qualidade dos dados e, conseqüentemente, nos sistemas de informação.

## REFERÊNCIAS

- AL-HAKIM, L. Information Quality Function Deployment. In: INTERNATIONAL CONFERENCE ON INFORMATION QUALITY (ICIQ 04), 9., p.170-182, 2004. **Proceedings ...** 2004.
- ALMEIDA, M.V.S. et al. Avaliação da qualidade dos dados do Sistema de Informação do Câncer do Colo do Útero em Vitória - ES, Brasil. **Revista Brasileira de Cancerologia**, v. 56, n. 3, p. 427 - 433. 2012.
- AMARAL, G.C.M. **AQUAWARE**: um ambiente de suporte à qualidade de dados em *Data Warehouse*. Dissertação (Mestrado em Informática)- Universidade Federal do Rio de Janeiro. Rio de Janeiro. 165p. 2003.
- BATINI, C. **Data Quality: Concepts, Methodologies and Techniques (Data-Centric System and Applications)**. Springer, New York, 1998.
- BATINI, C.; SCANNAPIECA, Monica. **Data Quality: Concepts, Methodologies and Techniques (Data-Centric System and Applications)**. Springer, New York, 1998.
- BELL, D. **The Coming of the Post-Industrial Society: A Venture in Social Forecasting**. New York: Basic Books, 1999.
- BOBROWSKI, Mónica; MARRÉ, Martina; YANKELEICH, Daniel. **Measuring Data Quality**. 1999. Relatório Técnico – Departamento de Computação – Faculdade de Ciências Exatas e Naturais, Universidade de Buenos Aires, Buenos Aires. Disponível em: <http://www.pragmaconsultores.com/uy/actualidad/Documents/Measuring%20Data%20Quality.PDF>. Acesso em: : 10 nov. 2013.
- BORLAND. **Delphi 7 Developer Guide**. Borland Software Corporation. 100 Enterprise Way, Scotts Valley. CA. 2002 Disponível em: [http://docs.embarcadero.com/products/rad\\_studio/delphi7/D7\\_DevelopersGuide.pdf](http://docs.embarcadero.com/products/rad_studio/delphi7/D7_DevelopersGuide.pdf) > Acesso em: 20 fev. 2014.
- BRANSKI, Regina Meyer; FRANCO, Raul Arellano Caldeira; LIMA JR, Orlando Fontes. Metodologia de estudo de casos aplicada à logística. In: CONGRESSO DE PESQUISA E ENSINO EM TRANSPORTES (XXIII ANPET), 25.. 2010. **Anais...** 2010.
- BUENO, Cassiane de Fátima dos Santos. **Qualidade de Software**. p. 1 - 28. Universidade Federal de Pernambuco - PE. Departamento de Informática. Disponível em <http://www.cin.ufpe.br/~mrsj/Qualidade/Qualidade%20de%20Software.pdf>. Acesso em: : 10 nov. 2013.
- BURBECK, S. **Applications programming in smalltalk-80 (tm): How to use model-view-controller (mvc)**. Smalltalk-80 v2. 5. ParcPlace, 80, 1–11. 1992. Disponível em [http://www.dgp.toronto.edu/~dwigdor/teaching/csc2524/2012\\_F/papers/mvc.pdf](http://www.dgp.toronto.edu/~dwigdor/teaching/csc2524/2012_F/papers/mvc.pdf) Acesso em: fev de 2014.
- CALAZANS, A.T.S.; COSTA, S.M.S. Modelo de avaliação da qualidade da informação estratégica bancária. **Ciência da Informação**. [online]. 2009, v.38, n.3, p. 21-39 . Disponível em: [www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19652009000300002&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652009000300002&lng=en&nrm=iso)>. Acesso em: : 10 jan. 2014.

CAVALCANTI, M.; GOMES, E. A Nova Riqueza das Organizações: Os Capitais do Conhecimento. **Revista Brasileira de Tecnologia e Negócios de Petróleo, Petroquímica, Química Fina, Gás e Indústria do Plástico**, Ano II, n. 16, 2000. Disponível em: <http://www.crie.coppe.ufrj.br>. Acesso em: : 23 set 2013.

DE PAULA, Larissa dos Santos. Análise da qualidade da informação em um processo de migração de dados. 133p. 2009. Dissertação (Mestrado em Engenharia de Produção e Sistema- Pontifícia Universidade Católica do Paraná. Curitiba. 2009.

DRUCKER, P. **Post-capitalist Society**. [S.l.]: , Butterworth-Heinemann. 1993.

ECKERSON, Wayne. W. **Data quality and the botttom line**: achieving business success through a commitment to high quality data. the data warehousing institute report series, USA. Disponível em: <http://download.101com.com/pub/tdwi/Files/DQReport.pdf>. Acesso em: nov 2013.

FAVARETTO, F. Proposta de medição da qualidade da informação. In: SIMPEP, 12., Bauru - SP, p. 1 - 7, 2005. **Anais...** 2005

FAVARETTO, F; VIEIRA, G.E. Estudo descritivo da qualidade da informação no planejamento da produção. **Revista Gestão Industrial**, v. 3, n. 2, p. 17 - 27. 2007

FAVARETTO, F. Experimento para análise da implantação da medição da qualidade da informação. **Associação Brasileira de Engenharia de Produção**, v.17 n.1 São Paulo jan./abr. 2007

GUEDES, G. T. A. **UML2**: uma abordagem prática. São Paulo: Novatec Editora Ltda., 2009

GUEDES, T. A. et al. **Projeto de ensino**: aprender fazendo estatística. 2012. Disponível em: <[http://www.uspleste.usp.br/rvicente/Estatistica\\_Descritiva.pdf](http://www.uspleste.usp.br/rvicente/Estatistica_Descritiva.pdf)>. Acesso em: 12 fev. 2014.

HEUSER, C. A. **Projeto de banco de dados**. 5. ed. [S.l.]: Instituto de Informática da UFRGS, Sagra-Luzzatto, 2000. (Série Livros Didáticos, 4).

ISO/IEC 9126-1: 2000. **Software engineering** - software product quality- Part 1: Quality Model. Disponível em: <[http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=22749](http://www.iso.org/iso/catalogue_detail.htm?csnumber=22749)> Acesso em: 12. jul 2013.

KOSCIANSKI, André; SOARES, Michel dos Santos. **Qualidade de software**: aprenda as metodologias e técnicas mais modernas para o desenvolvimento de software. 2. ed. São Paulo: Novatec. 395p. 2007.

KOSSWSKI, Lorete. A falta de qualidade de dados em sistemas organizacionais. **Revista Eletrônica Administração & Ciências Contábeis**, n.4, p.1 - 13, 2010.

KOVAC, R.L., Y.; PIPINO, L. **Total data quality management**: the case of IRI. *IQ*, 63–79. 1997. Disponível em: <[http://pdf.aminer.org/000/413/591/total\\_data\\_quality\\_management\\_the\\_case\\_of\\_iri.pdf](http://pdf.aminer.org/000/413/591/total_data_quality_management_the_case_of_iri.pdf)>. Acesso em: 12 out. 2013.



- LEE, Y.W. et al. AIMQ: methodology for information quality assessment. **Information & Management**, 2001
- LIMA, Claudia Risso de Araujo et al. Revisão das dimensões dos dados e métodos aplicados na avaliação dos sistemas de informação em saúde. **Cadernos de Saúde Pública**, n. 25, v. 10, p.2095 – 2109, 2009.
- MCGILVRAY, Danette. **Executing data quality projects: ten steps to quality data and trusted information**. [S.l.]: Morgan Kaufmann, 2008
- MICROSOFT CORPORATION, HYPERION SOLUTIONS CORPORATION. **XML for Anaysis Specification, v.1.1**. 2002. Disponível em: <[http://www.xmla.org/docs\\_pub.asp.2003](http://www.xmla.org/docs_pub.asp.2003)>. Acesso em: nov 2013
- MOSS, Larissa. Data Quality is not optimal. **Information Quality Conferency**, Denver, nov. 2003.
- O'BRIEN, James A. **Sistemas de informação e as decisões gerenciais na era da Internet**. 2. ed. São Paulo: Saraiva, 2004
- OBSERVATOIRE DES SCIENCES ET DESTECNIQUES. **Science & technologie; indicateurs 1998**. Paris: Economica, 1998.
- OLIVEIRA J.N.; AMARAL, L.A. O papel da qualidade da informação nos sistemas de informação. In: CONFERÊNCIA ESPECIALIZADA EM SISTEMAS E TECNOLOGIAS DE INFORMAÇÃO, 1999. **Anais...** [S.l.]: Universidade Católica Portuguesa. 1999. p. 1 - 17.
- OLSON, J. E. **Data quality: the accuracy dimension**. San Francisco, CA: Morgan Kaufmann Publishers, 2003.
- ORACLE® Database Concepts. Disponível em: <[http://docs.oracle.com/cd/B19306\\_01/server.102/b14220/triggers.htm](http://docs.oracle.com/cd/B19306_01/server.102/b14220/triggers.htm)> Acesso em: fev. 2014.
- PINHO, Selma, Foligne Crespio. **Avaliação da qualidade de dados pela não conformidade**. 2001. Dissertação (Mestrado em Ciências em Engenharia de Sistemas e Computação)- Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2001.
- PIPINO, Leo L.; LEE, Yang W.; WANG, Richard Y. Data Quality Assessment. **Communications of the ACM**. V. 45, n. 4, p. 211 - 218. 2002.
- PRESSMAN, Roger S. **Engenharia de software**. 6. ed. Porto Alegre: Bookman, 2006
- SALVADOR, V.F.M. et al. Qualidade de dados para gestão de conhecimento na área de Saúde. 2012. In: CBIS - CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE, 13., 2012. **Anais...** 2012. Disponível em: <http://www.sbis.org.br/cbis/arquivos/758.pdf>. Acesso em: nov de 2013.
- SILVA, E.O.; STANLEY, LOH. **Um software para avaliação de qualidade de fontes de informação em saúde na Web**. 2011. Disponível em: <[http://www.ulbra.inf.br/joomla/images/documentos/TCCs/2011\\_02/TCC\\_EDEMAR\\_OLIVEIRA.pdf](http://www.ulbra.inf.br/joomla/images/documentos/TCCs/2011_02/TCC_EDEMAR_OLIVEIRA.pdf)>. Acesso em: 10 jan. 2014.
- SIMÕES, R.P. **Informação, inteligência e utopia: contribuições à teoria de relações públicas**. [S.l.]: Summus, 2006.

SOMMERVILLE, Ian. Engenharia de software. 8. ed. São Paulo: Pearson Addison-Wesley, 2007

SOUZA, R.P.M. Análise da qualidade da informação em indicadores de desempenho utilizados em processos de gestão da cadeia de suprimentos: um estudo de caso. 2009. Dissertação (Mestrado em Engenharia de Produção e Sistemas)- Pontifícia Universidade Católica do Paraná – PUC/PR. Curitiba, 2009.

STOREY, V. C.; WANG, R. Y. An analysis of quality requirements in database design. In: INTERNATIONAL CONFERENCE ON INFORMATION QUALITY, 4., 1998. **Proc...** 1998.

TAYI, G. T., Ballou, D. P. Examining data quality. **Communications of the ACM**, v. 41, n. 2, 1998.

TORELLI, M. **Qualidade da informação sob a perspectiva do produto**. 76p. 2005. Monografia (Master Business Information Systems - MBIS Executivo em Ciência da Computação)- Pontifícia Universidade Católica de São Paulo - PUCSP, São Paulo, 2005

WAND, Y.; WANG, Richard. Y. Anchoring data quality dimensions in ontological foundations. **Communications of the ACM**, v.39, n. 11, p. 86 – 95, 1996.

WANG, Richard. Y. A Product Perspective on Total Data Quality Management. **Communications of the ACM**, V. 41, n. 2, 1998.

WANG, R.Y.; STRONG, D. M. Beyond Accuracy: what data quality means to data consumers. **Journal of Management Information Systems**, v.12, n.4, pp. 5-33, 1996.

WANG, R.Y.; ZIAD, M.; LEE, Y. W. **Data Quality**. [S.l.]: Kluwer Academic Publishers, 2000.

WANG, R.Y.; KON, H.B.; MADNICK, S. E. Data quality requirements analysis and modeling. In: INTERNATIONAL CONFERENCE OF DATA ENGINEERING, 9., 1993, Vienna, Austria. **Proc.** 1993.

WANG, R. Y.; REEDY, M. P.; KON, H. B. Toward quality data: an attribute-based approach. **Decision Support System**, v. 13, 1995

ZINS, C. Conceptual approaches for defining data, information, and knowledge. **Journal of the American Society for Information Science**, v.58, n.4, p.479-493, 2007.

## APÊNDICE A - DICIONÁRIO DE DADOS

### II. Dicionário de Dados da ferramenta

<b>Entidade</b>	TEMPLATES		
<b>Descrição:</b>	Tabela que armazena os templates dos gatilhos usados na ferramenta		
<b>Atributo:</b>	<b>Descrição</b>	<b>Tipo</b>	<b>Tamanho</b>
TEM_CODIGO	Chave primária da tabela	SMALLINT	
TEM_DESCRICA0	Código da trigger(gatilho)	BLOB	

<b>Entidade</b>	DIMENSOES		
<b>Descrição:</b>	Repositório das dimensões a serem utilizadas na ferramenta		
<b>Atributo:</b>	<b>Descrição</b>	<b>Tipo</b>	<b>Tamanho</b>
DIM_CODIGO	Chave primária da tabela	SMALLINT	
DIM_DESCRICA0	Nome da dimensão a ser trabalhada	VARCHAR	255

<b>Entidade</b>	PROJETOS		
<b>Descrição:</b>	Repositório onde são armazenados os dados referentes aos projetos testados pela ferramenta.		
<b>Atributo:</b>	<b>Descrição</b>	<b>Tipo</b>	<b>Tamanho</b>
PRO_CODIGO	Chave primária da tabela	SMALLINT	
PRO_DESCRICA0	Descrição do projeto	VARCHAR	255
PRO_CAMINHO	Caminho do banco de dados do projeto em teste	VARCHAR	512
PRO_DATA_INICIO	Data em que se iniciaram os testes	DATE	

<b>Entidade</b>	PROJETOS_DIMENSOES		
<b>Descrição:</b>	Tabela de agregação entre as tabelas PROJETOS e DIMENSOES		
<b>Atributo:</b>	<b>Descrição</b>	<b>Tipo</b>	<b>Tamanho</b>
PRD_CODIGO	Chave primária da tabela	SMALLINT	

PRD_DIM_CODIGO	Chave estrangeira da tabela DIMENSOES	SMALLINT	
PRD_PRO_CODIGO	Chave estrangeira da tabela PROJETOS	SMALLINT	
PRD_TRIGGER	Código do gatilho a ser inserido no sistema a ser testado	BLOB	
PRD_TIPO	Determina se o teste será automatizado ou manual	VARCHAR	20
PRD_TABELA	Nome da tabela a ser avaliada	VARCHAR	255
PRD_CAMPO	Nome do atributo a ser avaliado	VARCHAR	255
PRD_PESO	Peso para calculo da qualidade (opcional)	NUMERIC	12,2
PRD_BEFORE_AFTER	Determina se a trigger será BEFORE ou AFTER	VARCHAR	20
PRD_INSERT_UPDATE_DELETE	Determina se a trigger será para o INSERT, DELETE ou UPDATE	VARCHAR	100
PRD_DESCRICAÇÃO	Descrição da relação para nomeação do relatório	VARCHAR	255

### III. Dicionário de Dados do modelo alterado pela ferramenta

<b>Entidade</b>	TES\$DIMENSOES		
<b>Descrição:</b>	Repositório das dimensões a serem utilizadas na ferramenta		
<b>Atributo:</b>	<b>Descrição</b>	<b>po</b>	<b>Tamanho</b>
DIM_CODIGO	Chave primária da tabela	SMALLINT	
DIM_DESCRICAÇÃO	Nome da dimensão a ser trabalhada	VARCHAR	255

<b>Entidade</b>	TES\$AGREGACAO		
<b>Descrição:</b>	Agregação entre a tabela de dimensões e a de valores		
<b>Atributo:</b>	<b>Descrição</b>	<b>Tipo</b>	<b>Tamanho</b>
POS_CODIGO	Chave primária da tabela	SMALLINT	
POS_TABELA	Nome da tabela a ser avaliada	VARCHAR	255
POS_CAMPO	Nome do atributo a ser avaliado	VARCHAR	255
POS_PESO	Peso para calculo da qualidade (opcional)	NUMERIC	12,2
POS_DIM_CODIGO	Determina a chave estrangeira da dimensão a ser testada	SMALLINT	
POS_DATA_INICIO_TESTE	Data de Inicio dos Testes	DATE	
POS_DESCRICAÇÃO	Descrição da relação para nomeação do relatório	VARCHAR	255

<b>Entidade</b>	TES\$VALORES_FUNCIONARIOS		
<b>Descrição:</b>	Repositório onde são armazenados os valores referentes aos teste de QI realizados. Para cada tabela testada, uma entidade de nome TES\$VALORES_<NOME DA TABELA> será criada.		
<b>Atributo:</b>	<b>Descrição</b>	<b>Tipo</b>	<b>Tamanho</b>
VAL_CODIGO	Chave primária da tabela	SMALLINT	
VAL_POS_CODIGO	Chave estrangeira da tabela de agregação TES\$AGREGACAO	VARCHAR	255
VAL_FUN_CODIGO	Chave estrangeira da tabela em testes	VARCHAR	255
VAL_VALOR	Valor calculado para a qualidade do registro atual	NUMERIC	12,2
VAL_DESCRICAO	Descrição da qualidade no registro atual (opcional)	VARCHAR	255
VAL_DATA	Data em que o evento ocorreu	DATE	
VAL_HORA	Hora em que o evento ocorreu	TIME	255

## APÊNDICE B - TRIGGERS DE TESTES

### I. Triggers para o Cliente A

#### 1. Credibilidade da função de estorno da mensalidade

```

SET TERM ^ ;

CREATE TRIGGER TRIGGER3 FOR EVENTO
ACTIVE AFTER INSERT OR UPDATE
POSITION 100
AS
    DECLARE VARIABLE V_ETE_TEV_CODIGO VARCHAR(13);
    DECLARE VARIABLE V_ACA_VALIDADE DATE;
    DECLARE VARIABLE V_ACA_VALIDADE_PARA DATE;
    DECLARE VARIABLE V_POS_CODIGO INTEGER;
    DECLARE VARIABLE V_POS_PESO NUMERIC(18, 2);
BEGIN
    SELECT POS_CODIGO, POS_PESO
        FROM TES$AGREGACAO
        WHERE POS_TABELA = 'EVENTO'
            AND POS_CAMPO = 'EVE_NUMERO'
        INTO :V_POS_CODIGO, :V_POS_PESO;

    IF (NEW.EVE_SITUACAO = 'CANCELADA' ) THEN
    BEGIN

        SELECT ETE_TEV_CODIGO
            FROM EVENTO_TIPO_EVENTO
            WHERE ETE_TEV_CODIGO = '0001'
                AND ETE_EVE_NUMERO = NEW.EVE_NUMERO
        INTO :V_ETE_TEV_CODIGO;
    
```

```
IF (V_ETE_TEV_CODIGO IS NOT NULL) THEN
BEGIN

    SELECT ACA_VALIDADE, ACA_VALIDADE_PARA
    FROM ACADEMICO
    WHERE ACA_MATRICULA = NEW.EVE_ACA_MATRICULA
INTO :V_ACA_VALIDADE, :V_ACA_VALIDADE_PARA;
    V_ACA_VALIDADE = '1.1.2012';
    V_ACA_VALIDADE_PARA = '1.1.2013';
    IF (V_ACA_VALIDADE_PARA = V_ACA_VALIDADE ) THEN
        INSERT INTO TES$VALORES_EVENTO (VAL_CODIGO,
VAL_POS_CODIGO, VAL_EVE_NUMERO, VAL_VALOR)
            VALUES (GEN_ID(TES$VALORES_VAL_CODIGO_GEN,
1), :V_POS_CODIGO, NEW.EVE_NUMERO, 0);
        ELSE
            INSERT INTO TES$VALORES_EVENTO (VAL_CODIGO,
VAL_POS_CODIGO, VAL_EVE_NUMERO, VAL_VALOR)
            VALUES (GEN_ID(TES$VALORES_VAL_CODIGO_GEN,
1), :V_POS_CODIGO, NEW.EVE_NUMERO, 1);

    END

    END

    END
^

SET TERM ; ^
```

## 2. Credibilidade da função de estorno da mensalidade

```

SET TERM ^ ;

CREATE TRIGGER TRIGGER1 FOR PAGAMENTOS
ACTIVE AFTER INSERT OR UPDATE
POSITION 100
AS
DECLARE VARIABLE V_POS_CODIGO INTEGER;
DECLARE VARIABLE V_POS_PESO NUMERIC(18, 2);
BEGIN
    SELECT POS_CODIGO, POS_PESO
        FROM TES$AGREGACAO
        WHERE POS_TABELA = 'PAGAMENTOS'
            AND POS_CAMPO = 'PAG_TIPO_PAGAMENTO'
        INTO :V_POS_CODIGO, :V_POS_PESO;

    IF (NEW.PAG_TIPO_PAGAMENTO = 'CARTAO' ) THEN
    BEGIN
        IF (NEW.PAG_CAR_CODIGO IS NULL) THEN
            INSERT INTO TES$VALORES_PAGAMENTOS (VAL_CODIGO,
VAL_POS_CODIGO, VAL_PAG_NUMERO, VAL_VALOR)
                VALUES (GEN_ID(TES$VALORES_VAL_CODIGO_GEN,
1), :V_POS_CODIGO, NEW.PAG_NUMERO, 0);
        ELSE
            INSERT INTO TES$VALORES_PAGAMENTOS (VAL_CODIGO,
VAL_POS_CODIGO, VAL_PAG_NUMERO, VAL_VALOR)
                VALUES (GEN_ID(TES$VALORES_VAL_CODIGO_GEN,
1), :V_POS_CODIGO, NEW.PAG_NUMERO, :V_POS_PESO);

    END

```



```
END
```

```
^
```

```
SET TERM ; ^
```

### 3. Credibilidade da função de estorno da mensalidade

```
SET TERM ^ ;
```

```
CREATE TRIGGER TRIGGER2 FOR EVENTO
```

```
ACTIVE AFTER INSERT OR UPDATE
```

```
POSITION 100
```

```
AS
```

```
DECLARE VARIABLE V_POS_CODIGO INTEGER;
```

```
DECLARE VARIABLE V_POS_PESO NUMERIC(18, 2);
```

```
DECLARE VARIABLE V_USU_NIVEL VARCHAR(50);
```

```
BEGIN
```

```
    SELECT POS_CODIGO, POS_PESO
```

```
        FROM TESSAGREGACAO
```

```
        WHERE POS_TABELA = 'EVENTO'
```

```
            AND POS_CAMPO = 'EVE_SITUACAO'
```

```
    INTO :V_POS_CODIGO, :V_POS_PESO;
```

```
IF (NEW.EVE_SITUACAO = 'CANCELADA' ) THEN
```

```
BEGIN
```

```
    SELECT USU_NIVEL
```

```
        FROM USUARIO
```

```
        WHERE USU_CODIGO = NEW.EVE_USU_CODIGO_CANCELAMENTO
```

```
    INTO :V_USU_NIVEL;
```

```
IF (V_USU_NIVEL = 'ADMINISTRADOR') THEN
```

```
    INSERT INTO TESSVALORES_EVENTO (VAL_CODIGO,
```

```
VAL_POS_CODIGO, VAL_EVE_NUMERO, VAL_VALOR)
```

```
VALUES      (GEN_ID(TES$VALORES_VAL_CODIGO_GEN,  
1), :V_POS_CODIGO, NEW.EVE_NUMERO, :V_POS_PESO);  
ELSE  
INSERT      INTO      TES$VALORES_EVENTO (VAL_CODIGO,  
VAL_POS_CODIGO, VAL_EVE_NUMERO, VAL_VALOR)  
VALUES      (GEN_ID(TES$VALORES_VAL_CODIGO_GEN,  
1), :V_POS_CODIGO, NEW.EVE_NUMERO, 0);
```

```
END
```

```
END
```

```
^
```

```
SET TERM ; ^
```

## II. Triggers para o Cliente B

### 1. Completeza no Cancelamento de Aluguel

```

SET TERM ^ ;

CREATE TRIGGER TRIGGER1 FOR ALUGUEIS
ACTIVE AFTER INSERT OR UPDATE
POSITION 100
AS
DECLARE VARIABLE V_POS_CODIGO INTEGER;
DECLARE VARIABLE V_POS_PESO NUMERIC(18, 2);
DECLARE VARIABLE V_VALOR NUMERIC(18, 2);
BEGIN
    IF (NEW.ALU_SITUACAO = 'CANCELADO') THEN
        BEGIN
            SELECT POS_CODIGO, POS_PESO
            FROM TES$AGREGACAO
            WHERE POS_TABELA = 'ALUGUEIS'
            AND POS_CAMPO =
'ALU_FUN_CODIGO_CANCELAMENTO'
            INTO :V_POS_CODIGO, :V_POS_PESO;

            IF (NEW.ALU_FUN_CODIGO_CANCELAMENTO IS NULL)
THEN
                V_VALOR = 0;
            ELSE
                V_VALOR = 1;

            INSERT INTO TES$VALORES_ALUGUEIS (VAL_CODIGO,
VAL_POS_CODIGO, VAL_ALU_CODIGO, VAL_VALOR)

```

```
VALUES      (GEN_ID(TES$VALORES_VAL_CODIGO_GEN,  
1), :V_POS_CODIGO, NEW.ALU_CODIGO, :V_VALOR);  
      END  
END  
^  
  
SET TERM ; ^
```

## 2. Credibilidade no Lançamento de Telefonia

```

SET TERM ^ ;

CREATE TRIGGER TRIGGER2 FOR TELEFONEMAS
ACTIVE AFTER INSERT OR UPDATE
POSITION 100
AS
DECLARE VARIABLE V_POS_CODIGO INTEGER;
DECLARE VARIABLE V_POS_PESO NUMERIC(18, 2);
BEGIN
    SELECT POS_CODIGO, POS_PESO
        FROM TES$AGREGACAO
        WHERE POS_TABELA = 'TELEFONEMAS'
            AND POS_CAMPO = 'TEL_HAP_HOS_CODIGO'
        INTO :V_POS_CODIGO, :V_POS_PESO;
    IF (NEW.TEL_HAP_HOS_CODIGO IS NULL) THEN
        INSERT INTO TES$VALORES_TELEFONEMAS (VAL_CODIGO,
VAL_POS_CODIGO, VAL_TEL_CODIGO, VAL_VALOR)
            VALUES (GEN_ID(TES$VALORES_VAL_CODIGO_GEN,
1), :V_POS_CODIGO, NEW.TEL_CODIGO, 0);
    ELSE
        INSERT INTO TES$VALORES_TELEFONEMAS (VAL_CODIGO,
VAL_POS_CODIGO, VAL_TEL_CODIGO, VAL_VALOR)
            VALUES (GEN_ID(TES$VALORES_VAL_CODIGO_GEN,
1), :V_POS_CODIGO, NEW.TEL_CODIGO, :V_POS_PESO);

END
^

SET TERM ; ^

```

### 3. Temporalidade na hora de entrada na hospedagem do cliente

```

SET TERM ^ ;

CREATE TRIGGER TRIGGER4 FOR HOSPEDAGENS_APARTAMENTOS
ACTIVE AFTER INSERT OR UPDATE
POSITION 100
AS
DECLARE VARIABLE V_POS_CODIGO INTEGER;
DECLARE VARIABLE V_POS_PESO NUMERIC(18, 2);
BEGIN
    SELECT POS_CODIGO, POS_PESO
        FROM TES$AGREGACAO
        WHERE POS_TABELA = 'HOSPEDAGENS_APARTAMENTOS'
            AND POS_CAMPO = 'HAP_DATA_ENTRADA'
        INTO :V_POS_CODIGO, :V_POS_PESO;

    /* Sera considerado valido, clientes registrados ate 10
    minutos depois da entrada */
    IF      (NEW.HAP_DATA_ENTRADA      =      'TODAY'      AND
NEW.HAP_HORA_ENTRADA BETWEEN (CAST( 'NOW'AS TIME) - 100) AND
(CAST( 'NOW'AS TIME) + 600) ) THEN
        INSERT INTO
TES$VALORES_HOSPEDAGENS_APARTAM(VAL_CODIGO,      VAL_POS_CODIGO,
VAL_HAP_APA_CODIGO,      VAL_HAP_CLI_CODIGO,VAL_HAP_HOS_CODIGO,
VAL_VALOR)
            VALUES      (GEN_ID(TES$VALORES_VAL_CODIGO_GEN,
1),      :V_POS_CODIGO,      NEW.HAP_APA_CODIGO,NEW.HAP_CLI_CODIGO,
NEW.HAP_HOS_CODIGO,      :V_POS_PESO);
    ELSE
        INSERT INTO
TES$VALORES_HOSPEDAGENS_APARTAM(VAL_CODIGO,      VAL_POS_CODIGO,
VAL_HAP_APA_CODIGO,      VAL_HAP_CLI_CODIGO,VAL_HAP_HOS_CODIGO,
VAL_VALOR)

```

```
VALUES      (GEN_ID(TES$VALORES_VAL_CODIGO_GEN,  
1),      :V_POS_CODIGO,      NEW.HAP_APA_CODIGO,NEW.HAP_CLI_CODIGO,  
NEW.HAP_HOS_CODIGO, 0);
```

```
END
```

```
^
```

```
SET TERM ; ^
```

### III. Triggers para o Cliente C

#### 1. Completeza no Registro de Frequência dos Associados

```

SET TERM ^ ;

CREATE TRIGGER TRIGGER1 FOR FREQUENCIA
ACTIVE AFTER INSERT OR UPDATE
POSITION 100
AS
DECLARE VARIABLE V_POS_CODIGO INTEGER;
DECLARE VARIABLE V_POS_PESO NUMERIC(18, 2);
BEGIN
    SELECT POS_CODIGO, POS_PESO
        FROM TES$AGREGACAO
        WHERE POS_TABELA = 'FREQUENCIA'
            AND POS_CAMPO = 'FRE_HORA_SAIDA'
        INTO :V_POS_CODIGO, :V_POS_PESO;

    IF (INSERTING) THEN
        INSERT INTO TES$VALORES_FREQUENCIA (VAL_CODIGO,
            VAL_POS_CODIGO, VAL_FRE_DATA,
            VAL_FRE_HORA_ENTRADA, VAL_FRE_CONTROLE, VAL_VALOR)
            VALUES (GEN_ID(TES$VALORES_VAL_CODIGO_GEN, 1),
                :V_POS_CODIGO, NEW.FRE_DATA, NEW.FRE_HORA_ENTRADA,
                NEW.FRE_CONTROLE, 0);
    ELSE
        IF (UPDATING AND NEW.FRE_HORA_SAIDA IS NOT NULL) THEN
            BEGIN
                UPDATE TES$VALORES_FREQUENCIA SET VAL_VALOR =
                :V_POS_PESO
                WHERE VAL_FRE_DATA = NEW.FRE_DATA

```



```
AND VAL_FRE_HORA_ENTRADA =  
NEW.FRE_HORA_ENTRADA
```

```
AND VAL_FRE_CONTROLE = NEW.FRE_CONTROLE;
```

```
END
```

```
END
```

```
^
```

```
SET TERM ; ^
```

## 2. Credibilidade no Registro de Entrada de Sócios

```

SET TERM ^ ;

CREATE TRIGGER TRIGGER2 FOR FREQUENCIA
ACTIVE AFTER INSERT
POSITION 100
AS
DECLARE VARIABLE V_POS_CODIGO INTEGER;
DECLARE VARIABLE V_POS_PESO NUMERIC(18, 2);
BEGIN
    SELECT POS_CODIGO, POS_PESO
        FROM TES$AGREGACAO
        WHERE POS_TABELA = 'FREQUENCIA'
            AND POS_CAMPO = 'FRE_OCORRENCIA_ENTRADA'
        INTO :V_POS_CODIGO, :V_POS_PESO;

    IF (NEW.FRE_OCORRENCIA_ENTRADA IS NOT NULL) THEN
        BEGIN
            IF (NEW.FRE_FUN_MATRICULA_LIBERACAO_ENT IS NULL) THEN
                INSERT INTO TES$VALORES_FREQUENCIA (VAL_CODIGO,
VAL_POS_CODIGO, VAL_FRE_DATA,
VAL_FRE_HORA_ENTRADA, VAL_FRE_CONTROLE, VAL_VALOR)
                    VALUES (GEN_ID(TES$VALORES_VAL_CODIGO_GEN,
1), :V_POS_CODIGO, NEW.FRE_DATA, NEW.FRE_HORA_ENTRADA,
NEW.FRE_CONTROLE, 0);
            ELSE
                INSERT INTO TES$VALORES_FREQUENCIA (VAL_CODIGO,
VAL_POS_CODIGO, VAL_FRE_DATA,
VAL_FRE_HORA_ENTRADA, VAL_FRE_CONTROLE, VAL_VALOR)
                    VALUES (GEN_ID(TES$VALORES_VAL_CODIGO_GEN,
1), :V_POS_CODIGO, NEW.FRE_DATA, NEW.FRE_HORA_ENTRADA,
NEW.FRE_CONTROLE, :V_POS_PESO);
        END
    END

```

END

END

^

SET TERM ; ^

### 3. Temporalidade no Registro de Mensalidades

```

SET TERM ^ ;

CREATE TRIGGER TRIGGER3 FOR MENSALIDADE
ACTIVE AFTER INSERT
POSITION 100
AS
DECLARE VARIABLE V_POS_CODIGO INTEGER;
DECLARE VARIABLE V_POS_PESO NUMERIC(18, 2);
DECLARE VARIABLE V_MES_LANCAMENTO INTEGER;
DECLARE VARIABLE V_ANO_LANCAMENTO INTEGER;
DECLARE VARIABLE V_MES INTEGER;
DECLARE VARIABLE V_ANO INTEGER;
BEGIN
    SELECT POS_CODIGO, POS_PESO
           FROM TES$AGREGACAO
           WHERE POS_TABELA = 'MENSALIDADE'
              AND POS_CAMPO = 'MEN_LANCAMENTO'
           INTO :V_POS_CODIGO, :V_POS_PESO;

    V_MES_LANCAMENTO = CAST(EXTRACT (MONTH FROM
NEW.MEN_LANCAMENTO) AS INTEGER);
    V_ANO_LANCAMENTO = CAST(EXTRACT (YEAR FROM
NEW.MEN_LANCAMENTO) AS INTEGER);
    V_MES = CAST(SUBSTRING(NEW.MEN_MES_ANO FROM 1 FOR 2) AS
INTEGER);
    V_ANO = CAST(SUBSTRING(NEW.MEN_MES_ANO FROM 3 FOR 4) AS
INTEGER);

    IF (V_MES = 1) THEN

```

```

BEGIN
    V_MES = 12;
    V_ANO = V_ANO - 1;
END
ELSE
    V_MES = V_MES - 1;

    IF (V_ANO = V_ANO_LANCAMENTO AND V_MES >=
V_MES_LANCAMENTO) THEN
        INSERT INTO TES$VALORES_MENSALIDADE (VAL_CODIGO,
VAL_POS_CODIGO, VAL_MEN_MES_ANO, VAL_MEN_SOC_MATRICULA,
VAL_VALOR)
            VALUES (GEN_ID(TES$VALORES_VAL_CODIGO_GEN, 1),
:V_POS_CODIGO, NEW.MEN_MES_ANO,NEW.MEN_SOC_MATRICULA,
:V_POS_PESO);
        ELSE
            INSERT INTO TES$VALORES_MENSALIDADE (VAL_CODIGO,
VAL_POS_CODIGO, VAL_MEN_MES_ANO, VAL_MEN_SOC_MATRICULA,
VAL_VALOR)
                VALUES (GEN_ID(TES$VALORES_VAL_CODIGO_GEN, 1),
:V_POS_CODIGO, NEW.MEN_MES_ANO,NEW.MEN_SOC_MATRICULA, 0);
        WHEN ANY DO
            BEGIN
                EXIT;
            END
        END

END
^

SET TERM ; ^

```

#### 4. Quantidade Adequada de Dados Na Função de Lançamento dos Boletos

```

SET TERM ^ ;

CREATE TRIGGER TRIGGER4 FOR BOLETO_MENSALIDADES
ACTIVE AFTER INSERT
POSITION 100
AS
DECLARE VARIABLE V_POS_CODIGO INTEGER;
DECLARE VARIABLE V_POS_PESO NUMERIC(18, 2);
DECLARE VARIABLE V_CONT INTEGER;
DECLARE VARIABLE V_EXISTE INTEGER;
BEGIN
    SELECT POS_CODIGO, POS_PESO
           FROM TES$AGREGACAO
           WHERE POS_TABELA = 'BOLETO_MENSALIDADES'
                  AND POS_CAMPO = 'BME_MEN_MES_ANO'
           INTO :V_POS_CODIGO, :V_POS_PESO;

    SELECT COALESCE(COUNT (*), 0)
           FROM BOLETO_MENSALIDADES
           WHERE BME_MEN_MES_ANO = NEW.BME_MEN_MES_ANO
                  AND
                  BME_MEN_SOC_MATRICULA
NEW.BME_MEN_SOC_MATRICULA
           INTO :V_CONT;

    SELECT COALESCE(MAX(VAL_CODIGO), 0)
           FROM TES$VALORES_BOLETO_MENSALIDADES
           WHERE VAL_BME_MEN_MES_ANO = NEW.BME_MEN_MES_ANO
                  AND
                  VAL_BME_MEN_SOC_MATRICULA
NEW.BME_MEN_SOC_MATRICULA
                  AND VAL_BME_BOL_CODIGO = NEW.BME_BOL_CODIGO

```

```

        INTO V_EXISTE;

        /* Foi definido pelo usuario 2 boletos como o maximo
para a questao da qualidade */
        IF (V_CONT > 2 ) THEN
        BEGIN
            IF (V_EXISTE = 0) THEN
                INSERT INTO
TES$VALORES_BOLETO_MENSALIDADES (VAL_CODIGO, VAL_POS_CODIGO,
VAL_BME_MEN_MES_ANO, VAL_BME_MEN_SOC_MATRICULA,
VAL_BME_BOL_CODIGO, VAL_VALOR)
                VALUES (GEN_ID(TES$VALORES_VAL_CODIGO_GEN,
1), :V_POS_CODIGO, NEW.BME_MEN_MES_ANO,
NEW.BME_MEN_SOC_MATRICULA, NEW.BME_BOL_CODIGO, 0);
            ELSE
                UPDATE TES$VALORES_BOLETO_MENSALIDADES
                SET VAL_VALOR = 0
                WHERE VAL_CODIGO = :V_EXISTE;

        END
        ELSE
            IF (V_EXISTE = 0) THEN
                INSERT INTO
TES$VALORES_BOLETO_MENSALIDADES (VAL_CODIGO, VAL_POS_CODIGO,
VAL_BME_MEN_MES_ANO, VAL_BME_MEN_SOC_MATRICULA,
VAL_BME_BOL_CODIGO, VAL_VALOR)
                VALUES (GEN_ID(TES$VALORES_VAL_CODIGO_GEN, 1),
:V_POS_CODIGO, NEW.BME_MEN_MES_ANO, NEW.BME_MEN_SOC_MATRICULA,
NEW.BME_BOL_CODIGO, :V_POS_PESO);
            ELSE
                UPDATE TES$VALORES_BOLETO_MENSALIDADES
                SET VAL_VALOR = :V_POS_PESO
                WHERE VAL_CODIGO = :V_EXISTE;

```

END

^

SET TERM ; ^



## **APÊNDICE C - QUESTIONARIO DE ENTREVISTA DE CLIENTE**

### **Apresentação**

Com o intuito de realizarmos uma pesquisa acerca da qualidade de dados em sua empresa, com a finalidade de embasar uma dissertação de mestrado que tem como tema a medição da qualidade de dados em ambientes transacionais, estamos apresentando o questionário abaixo.

A sua ajuda será muito importante para a nossa pesquisa e contribuirá para a melhoria da qualidade da informação em seu sistema, haja vista que os resultados lhe serão retornados para que seja traçado um plano de ação para a resolução dos problemas de qualidade em seu sistema.

Para compreender os fundamentos que embasam a ideia de qualidade de dados, é necessário definir, em primeira instância, o que se entende por qualidade. De acordo com Kossowski (2010), são diversos os conceitos sobre qualidade, expressão utilizada desde muito tempo, porém, tratando especificamente de qualidade em produtos. Evans & Lindsay (1999) consideram a qualidade um conjunto de traços e atributos de determinado produto ou serviço que busca alcançar ou exceder as expectativas e necessidades do cliente. Pinho (2001) considera que a obtenção da qualidade exige mudança de postura e paradigmas dentro de uma empresa, tendo sempre como objetivo a satisfação do cliente. Tal mudança de postura, ainda de acordo com a autora, envolve a percepção de tratar as respostas aos problemas como uma oportunidade de alcançar a confiança e satisfação do cliente, o que pode garantir, em última instância, um aumento na participação do mercado, bem como a fidelidade dos clientes. Percebe-se, portanto, que apesar de ter um conceito complexo, é consenso que a qualidade busca a satisfação de todos os indivíduos

envolvidos no processo, desde a empresa, acionistas e fornecedores, até o cliente.

Oliveira & Amaral (1999) discutem um aspecto importante da qualidade, que é o fato de que a qualidade não está somente nas características intrínsecas dos objetos, e sim na utilização ou aplicação desses. Deming apud Kossowski (2010), um especialista no assunto, considera que a qualidade é multidimensional e dinâmica, assim como as necessidades do mercado consumidor, e se concentra em diversos níveis, vez que um determinado produto pode satisfazer a dois clientes de diferentes maneiras.

A complexidade de definir qualidade está justamente no fato de que a qualidade vai possuir significados de acordo com as necessidades das diferentes pessoas e depende ainda da perspectiva do avaliador, o que torna complexa a atividade de construir uma medida de qualidade aceitável para todos (PINHO, 2001).

Não é necessário identificar-se.

Muito obrigado por sua colaboração!

Empresa(Opcional):	
Ramo de Atividade:	
Clientes Cadastrados:	Rotatividade (Mês):
Entrevistado(Opcional):	
Cargo:	

Baseado nos conceitos de qualidade apresentados, cite, ao menos, 3 (três) pontos críticos em seu sistema que você gostaria de ter a qualidade de dados medidas. Em seguida, classifique o grau de importância para o ponto citado, colocando 1(um) para **baixa**, 2(dois) para **média** ou 3(três) para **alta**. Seja bastante claro e objetivo na descrição dos pontos.

Ponto 1	Importância

Ponto 2	Importância

Ponto 3	Importância

Ponto 4	Importância

Ponto 5	Importância

## APÊNDICE D - SCRIPTS PARA A CRIAÇÃO DAS TABELAS DE DIMENSÕES

### I. Tabelas Padrões De Dimensão

```
/* Criacao da tabela TES$DIMENSOES */
```

```
CREATE TABLE TES$DIMENSOES (  
    DIM_CODIGO SMALLINT NOT NULL,  
    DIM_DESCRICAO VARCHAR(255);
```

```
/* Definicao da chave primaria da tabela TES$DIMENSOES */
```

```
ALTER TABLE TES$DIMENSOES ADD PRIMARY KEY (DIM_CODIGO);
```

```
/* Criacao da tabela TES$AGREGACAO */
```

```
CREATE TABLE TES$AGREGACAO (  
    POS_CODIGO SMALLINT NOT NULL,  
    POS_TABELA VARCHAR(255),  
    POS_CAMPO VARCHAR(255),  
    POS_PESO NUMERIC(12, 2) NOT NULL,  
    POS_DIM_CODIGO SMALLINT DEFAULT 0 NOT NULL,  
    POS_DESCRICAO VARCHAR(255),  
    POS_DATA_INICIO_TESTE DATE DEFAULT 'TODAY' NOT NULL);
```

```
/* Definicao da chave primaria da tabela TES$AGREGACAO */
```

```
ALTER TABLE TES$AGREGACAO ADD PRIMARY KEY (POS_CODIGO);
```

```
/* Criacao da tabela TES$VALORES_ FUNCIONARIOS */
```

```
CREATE TABLE TES$VALORES_ FUNCIONARIOS (  
    VAL_CODIGO INTEGER DEFAULT 0 NOT NULL,  
    VAL_POS_CODIGO SMALLINT NOT NULL,
```

```

VAL_FUN_CODIGO INTEGER NOT NULL,
VAL_VALOR NUMERIC(12, 2),
VAL_DESCRICAO VARCHAR(255),
VAL_HORA TIME DEFAULT NOW NOT NULL,
VAL_DATA DATE DEFAULT TODAY NOT NULL);

```

```

/* Definicao da chave primaria da tabela TES$VALORES_FUNCIONARIOS */

```

```

ALTER TABLE TES$VALORES_FUNCIONARIOS ADD PRIMARY KEY (VAL_CODIGO);

```

## II. Script para a Criação de Tabelas de Dimensão para o Cliente A

```

/* Criacao da tabela TES$VALORES_PAGAMENTOS */

```

```

CREATE TABLE TES$VALORES_PAGAMENTOS (
  VAL_CODIGO INTEGER DEFAULT 0 NOT NULL,
  VAL_POS_CODIGO SMALLINT NOT NULL,
  VAL_PAG_NUMERO INTEGER NOT NULL,
  VAL_VALOR NUMERIC(12, 2),
  VAL_DESCRICAO VARCHAR(255) ,
  VAL_HORA TIME DEFAULT NOW NOT NULL,
  VAL_DATA DATE DEFAULT TODAY NOT NULL,
  VAL_PAG_DUP_DATA DATE,
  VAL_PAG_DUP_CODIGO SMALLINT);

```

```

/* Definicao da chave primaria da tabela TES$VALORES_PAGAMENTOS */

```

```

ALTER TABLE TES$VALORES_PAGAMENTOS ADD PRIMARY KEY (VAL_CODIGO);

```

```

/* Definicoes das chaves estrangeiras da tabela
TES$VALORES_PAGAMENTOS */

```

```

ALTER TABLE TES$VALORES_PAGAMENTOS ADD CONSTRAINT
FK_TES$VALORES_PAGAMENTOS1 FOREIGN KEY
(VAL_PAG_DUP_CODIGO, VAL_PAG_DUP_DATA, VAL_PAG_NUMERO)
REFERENCES
PAGAMENTOS (PAG_DUP_CODIGO, PAG_DUP_DATA, PAG_NUMERO) ON UPDATE CASCADE;

```

```
ALTER TABLE TESSVALORES_PAGAMENTOS ADD CONSTRAINT
FK_TESSVALORES_PAGAMENTOS FOREIGN KEY (VAL_POS_CODIGO) REFERENCES
TESSAGREGACAO (POS_CODIGO) ON UPDATE CASCADE;
```

```
/* Criacao da tabela TESSVALORES_EVENTO*/
```

```
CREATE TABLE TESSVALORES_EVENTO (
  VAL_CODIGO INTEGER DEFAULT 0 NOT NULL,
  VAL_POS_CODIGO SMALLINT NOT NULL,
  VAL_EVE_NUMERO INTEGER NOT NULL,
  VAL_VALOR NUMERIC(12, 2),
  VAL_DESCRICAO VARCHAR(255) CHARACTER SET ASCII COLLATE ASCII,
  VAL_HORA TIME DEFAULT NOW NOT NULL,
  VAL_DATA DATE DEFAULT TODAY NOT NULL);
```

```
/* Definicao da chave primaria da tabela TESSVALORES_EVENTO */
```

```
ALTER TABLE TESSVALORES_EVENTO ADD PRIMARY KEY (VAL_CODIGO);
```

```
/* Definicoes das chaves estrangeiras da tabela TESSVALORES_EVENTO */
```

```
ALTER TABLE TESSVALORES_EVENTO ADD CONSTRAINT FK_TESSVALORES_EVENTO
FOREIGN KEY (VAL_POS_CODIGO) REFERENCES TESSAGREGACAO (POS_CODIGO) ON UPDATE
CASCADE;
```

```
ALTER TABLE TESSVALORES_EVENTO ADD CONSTRAINT FK_TESSVALORES_EVENTO1
FOREIGN KEY (VAL_EVE_NUMERO) REFERENCES EVENTO (EVE_NUMERO) ON UPDATE
CASCADE;
```

### III. Script para a Criação de Tabelas de Dimensão para o Cliente B

```
/* Criacao da tabela TESSVALORES_ALUGUEIS */
```

```
CREATE TABLE TESSVALORES_ALUGUEIS (
  VAL_CODIGO INTEGER DEFAULT 0 NOT NULL,
  VAL_POS_CODIGO SMALLINT NOT NULL,
  VAL_ALU_CODIGO INTEGER NOT NULL,
```

```

VAL_VALOR NUMERIC(12, 2),
VAL_DESCRICAO VARCHAR(255),
VAL_HORA TIME DEFAULT NOW NOT NULL,
VAL_DATA DATE DEFAULT TODAY NOT NULL);

```

```
/* Definicao da chave primaria da tabela TES$VALORES_ALUGUEIS */
```

```
ALTER TABLE TES$VALORES_ALUGUEIS ADD PRIMARY KEY (VAL_CODIGO);
```

```
/* Definicao das chaves estrangeiras da tabela TES$VALORES_ALUGUEIS*/
```

```

ALTER      TABLE      TES$VALORES_ALUGUEIS      ADD      CONSTRAINT
FK_TES$VALORES_ALUGUEIS      FOREIGN      KEY      (VAL_ALU_CODIGO)      REFERENCES
ALUGUEIS(ALU_CODIGO) ON UPDATE CASCADE;

```

```

ALTER      TABLE      TES$VALORES_ALUGUEIS      ADD      CONSTRAINT
FK_TES$VALORES_ALUGUEIS1      FOREIGN      KEY      (VAL_POS_CODIGO)      REFERENCES
TES$AGREGACAO(POS_CODIGO) ON UPDATE CASCADE;

```

```
/* Criacao da tabela TES$VALORES_HOSPEDAGENS_APARTAM */
```

```

CREATE TABLE TES$VALORES_HOSPEDAGENS_APARTAM (
  VAL_CODIGO INTEGER DEFAULT 0 NOT NULL,
  VAL_POS_CODIGO SMALLINT NOT NULL,
  VAL_HAP_APA_CODIGO INTEGER NOT NULL,
  VAL_VALOR NUMERIC(12, 2),
  VAL_DESCRICAO VARCHAR(255) CHARACTER SET ASCII COLLATE ASCII,
  VAL_HORA TIME DEFAULT NOW NOT NULL,
  VAL_DATA DATE DEFAULT TODAY NOT NULL,
  VAL_HAP_HOS_CODIGO INTEGER DEFAULT 0 NOT NULL,
  VAL_HAP_CLI_CODIGO INTEGER DEFAULT 0 NOT NULL);

```

```
/*      Definicao      da      chave      primaria      da      tabela
TES$VALORES_HOSPEDAGENS_APARTAM */
```

```

ALTER      TABLE      TES$VALORES_HOSPEDAGENS_APARTAM      ADD      PRIMARY      KEY
(VAL_CODIGO);

```

```
/*      Definicao      das      chaves      estrangeiras      da      tabela
TES$VALORES_HOSPEDAGENS_APARTAM */
```



```
ALTER TABLE TES$VALORES_HOSPEDAGENS_APARTAM ADD CONSTRAINT
FK_TES$VALORES_HOSPEDAGENS_AP2 FOREIGN KEY (VAL_POS_CODIGO) REFERENCES
TES$AGREGACAO (POS_CODIGO) ON UPDATE CASCADE;
```

```
ALTER TABLE TES$VALORES_HOSPEDAGENS_APARTAM ADD CONSTRAINT
FK_TES$VALORES_HOSPEDAGENS_APA2 FOREIGN KEY
(VAL_HAP_APA_CODIGO, VAL_HAP_CLI_CODIGO, VAL_HAP_HOS_CODIGO) REFERENCES
HOSPEDAGENS_APARTAMENTOS (HAP_APA_CODIGO, HAP_CLI_CODIGO, HAP_HOS_CODIGO) ON
UPDATE CASCADE;
```

```
/* Criacao da tabela TES$VALORES_TELEFONEMAS */
```

```
CREATE TABLE TES$VALORES_TELEFONEMAS (
VAL_CODIGO INTEGER DEFAULT 0 NOT NULL,
VAL_POS_CODIGO SMALLINT NOT NULL,
VAL_TEL_CODIGO INTEGER NOT NULL,
VAL_VALOR NUMERIC(12, 2),
VAL_DESCRICAO VARCHAR(255) CHARACTER SET ASCII COLLATE ASCII,
VAL_HORA TIME DEFAULT NOW NOT NULL,
VAL_DATA DATE DEFAULT TODAY NOT NULL);
```

```
Definicao da chave primaria da tabela TES$VALORES_TELEFONEMAS */
```

```
ALTER TABLE TES$VALORES_TELEFONEMAS ADD PRIMARY KEY (VAL_CODIGO);
```

```
/* Definicao das chaves estrangeiras da tabela
TES$VALORES_TELEFONEMAS */
```

```
ALTER TABLE TES$VALORES_TELEFONEMAS ADD CONSTRAINT
FK_TES$VALORES_TELEFONEMAS FOREIGN KEY (VAL_POS_CODIGO) REFERENCES
TES$AGREGACAO (POS_CODIGO) ON UPDATE CASCADE;
```

```
ALTER TABLE TES$VALORES_TELEFONEMAS ADD CONSTRAINT
FK_TES$VALORES_TELEFONEMAS1 FOREIGN KEY (VAL_TEL_CODIGO) REFERENCES
TELEFONEMAS (TEL_CODIGO) ON UPDATE CASCADE;
```

#### IV. Script para a Criação de Tabelas de Dimensão para o Cliente C

```
/* Criacao da tabela TES$VALORES_FREQUENCIA */
```

```
CREATE TABLE TES$VALORES_FREQUENCIA (
  VAL_CODIGO INTEGER DEFAULT 0 NOT NULL,
  VAL_POS_CODIGO SMALLINT NOT NULL,
  VAL_FRE_DATA DATE NOT NULL,
  VAL_VALOR NUMERIC(12, 2),
  VAL_DESCRICAO VARCHAR(255) CHARACTER SET ASCII COLLATE ASCII,
  VAL_HORA TIME DEFAULT NOW NOT NULL,
  VAL_DATA DATE DEFAULT TODAY NOT NULL,
  VAL_FRE_HORA_ENTRADA TIME DEFAULT 00:00:00 NOT NULL,
  VAL_FRE_CONTROLE INTEGER DEFAULT 0 NOT NULL);
```

```
/* Definicao da chave primaria da tabela TES$VALORES_FREQUENCIA */
```

```
ALTER TABLE TES$VALORES_FREQUENCIA ADD PRIMARY KEY (VAL_CODIGO);
```

```
/* Definicao das chaves estrangeiras da tabela TES$VALORES_FREQUENCIA
*/
```

```
ALTER TABLE TES$VALORES_FREQUENCIA ADD CONSTRAINT
FK_TES$VALORES_FREQUENCIA FOREIGN KEY (VAL_POS_CODIGO) REFERENCES
TES$AGREGACAO (POS_CODIGO) ON UPDATE CASCADE;
```

```
ALTER TABLE TES$VALORES_FREQUENCIA ADD CONSTRAINT
FK_TES$VALORES_FREQUENCIA1 FOREIGN KEY
(VAL_FRE_DATA, VAL_FRE_HORA_ENTRADA, VAL_FRE_CONTROLE) REFERENCES
FREQUENCIA (FRE_DATA, FRE_HORA_ENTRADA, FRE_CONTROLE) ON UPDATE CASCADE;
```

```
/* Criacao da tabela TES$VALORES_BOLETO_MENSALIDADES */
```

```
CREATE TABLE TES$VALORES_BOLETO_MENSALIDADES (
  VAL_CODIGO INTEGER DEFAULT 0 NOT NULL,
  VAL_POS_CODIGO SMALLINT NOT NULL,
  VAL_BME_MEN_MES_ANO VARCHAR(6) CHARACTER SET NONE NOT NULL COLLATE
NONE,
  VAL_BME_MEN_SOC_MATRICULA VARCHAR(15) CHARACTER SET NONE NOT NULL
COLLATE NONE,
```

```

VAL_BME_BOL_CODIGO VARCHAR(60) CHARACTER SET NONE NOT NULL COLLATE
NONE,
VAL_VALOR NUMERIC(12, 2),
VAL_DESCRICAO VARCHAR(255) CHARACTER SET ASCII COLLATE ASCII,
VAL_HORA TIME DEFAULT NOW NOT NULL,
VAL_DATA DATE DEFAULT TODAY NOT NULL);

```

```

/* Definicao da chave primaria da tabela
TES$VALORES_BOLETO_MENSALIDADES */

```

```

ALTER TABLE TES$VALORES_BOLETO_MENSALIDADES ADD PRIMARY KEY
(VAL_CODIGO);

```

```

/* Definicao das chaves estrangeiras da tabela
TES$VALORES_BOLETO_MENSALIDADES */

```

```

ALTER TABLE TES$VALORES_BOLETO_MENSALIDADES ADD CONSTRAINT
FK_TES$VALORES_BOLETO_MENSALI1 FOREIGN KEY (VAL_POS_CODIGO) REFERENCES
TES$AGREGACAO(POS_CODIGO) ON UPDATE CASCADE;

```

```

ALTER TABLE TES$VALORES_BOLETO_MENSALIDADES ADD CONSTRAINT
FK_TES$VALORES_BOLETO_MENSALI2 FOREIGN KEY
(VAL_BME_MEN_MES_ANO,VAL_BME_MEN_SOC_MATRICULA,VAL_BME_BOL_CODIGO)
REFERENCES
BOLETO_MENSALIDADES(BME_MEN_MES_ANO,BME_MEN_SOC_MATRICULA,BME_BOL_CODIGO)
ON UPDATE CASCADE;

```

```

/* Criacao da tabela TES$VALORES_MENSALIDADE */

```

```

CREATE TABLE TES$VALORES_MENSALIDADE (
VAL_CODIGO INTEGER DEFAULT 0 NOT NULL,
VAL_POS_CODIGO SMALLINT NOT NULL,
VAL_MEN_MES_ANO VARCHAR(6) CHARACTER SET NONE NOT NULL COLLATE
NONE,
VAL_VALOR NUMERIC(12, 2),
VAL_DESCRICAO VARCHAR(255) CHARACTER SET ASCII COLLATE ASCII,
VAL_HORA TIME DEFAULT NOW NOT NULL,
VAL_DATA DATE DEFAULT TODAY NOT NULL,

```

```
VAL_MEN_SOC_MATRICULA VARCHAR(15) CHARACTER SET ASCII NOT NULL  
COLLATE ASCII);
```

```
/* Definicao da chave primaria da tabela TES$VALORES_MENSALIDAD */
```

```
ALTER TABLE TES$VALORES_MENSALIDADE ADD PRIMARY KEY (VAL_CODIGO);
```

```
/* Definicao das chaves estrangeiras da tabela  
TES$VALORES_MENSALIDADE */
```

```
ALTER TABLE TES$VALORES_MENSALIDADE ADD CONSTRAINT  
FK_TES$VALORES_MENSALIDAD1 FOREIGN KEY (VAL_POS_CODIGO) REFERENCES  
TES$AGREGACAO (POS_CODIGO) ON UPDATE CASCADE;
```

```
ALTER TABLE TES$VALORES_MENSALIDADE ADD CONSTRAINT  
FK_TES$VALORES_MENSALIDADE FOREIGN KEY  
(VAL_MEN_MES_ANO, VAL_MEN_SOC_MATRICULA) REFERENCES  
MENSALIDADE (MEN_MES_ANO, MEN_SOC_MATRICULA) ON UPDATE CASCADE;
```