

EDSON YUKIO HOJI

**MELHORIA DE UM SISTEMA DE *DATA WAREHOUSE* EM UMA
EMPRESA DE TELEFONIA MÓVEL**

**Trabalho de formatura apresentado à
Escola Politécnica da Universidade de
São Paulo para obtenção do Diploma
de Engenheiro de Produção**

São Paulo

2012

EDSON YUKIO HOJI

**MELHORIA DE UM SISTEMA DE *DATA WAREHOUSE* EM UMA
EMPRESA DE TELEFONIA MÓVEL**

**Trabalho de formatura apresentado à
Escola Politécnica da Universidade de
São Paulo para obtenção do Diploma
de Engenheiro de Produção**

**Orientador: Prof. Livre-Docente
Marcelo Schneck de Paula Pessoa**

São Paulo

2012

FICHA CATALOGRÁFICA

Hoji, Edson Yukio

Melhoria de um sistema de Data Warehouse em uma empresa de telefonia móvel / E.Y. Hoji. -- São Paulo, 2012.

95 p.

Trabalho de Formatura - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Produção.

**1. Tecnologia da informação 2. Sistemas de informação
3. Banco de dados 4. Telefonia móvel I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Produção II. t.**

DEDICATÓRIA

Dedico este trabalho aos meus pais, Masakazu e Isaura, por toda a educação recebida deles e por servirem de exemplo como pessoas, e à minha namorada Luciana por toda paciência, suporte e carinho durante a elaboração deste trabalho.

AGRADECIMENTOS

A minha família, por sempre me apoiar e fornecer o ambiente propício para o meu crescimento pessoal e profissional, além de mostrar que o caminho do estudo é recompensador. Agradeço também por terem me dado a oportunidade de estudar em boas escolas e bons colégios.

Ao Professor Marcelo Pessôa, pela paciência, pelos conselhos e pela orientação fundamentais durante a elaboração deste trabalho.

A Luciana pelo carinho, amor, apoio e paciência.

A todos os colegas de trabalho que me auxiliaram e me apoiaram durante o curto período que passamos juntos.

RESUMO

As empresas necessitam de sistemas capazes de tratar a enorme quantidade de dados gerados nos dias atuais. É necessário um sistema que não somente registre os dados operacionais, mas que também os armazene e os disponibilize para consultas, de forma organizada. Esse sistema é conhecido por *data warehouse*, e trata os dados de maneira a torna-los úteis no processo de tomada de tomada de decisão. Assim, o objetivo deste trabalho visa apresentar um projeto de melhoria de um sistema de *data warehouse* em uma empresa de telefonia móvel. Com mudanças de necessidades e requisitos, foi necessária a readequação da solução implantada de *data warehouse*. Como principais resultados do projeto, estão a aderência do sistema às necessidades dos usuários, a crescente utilização do sistema e o aumento de sua confiabilidade.

Palavras-chave: *Data warehouse*. Modelagem Dimensional. Sistemas de apoio a decisão. Telefonia Móvel. Sistemas de Informação.

ABSTRACT

Nowadays, companies need systems capable of handling the huge amount of data that is generated. It is needed a system not only for recording operational data, but also able to store and make them available for queries, in an organized fashion. This system is known as data warehouse, and handles the data in a way that make them useful for decision support. The objective of this paper is to present an enhancement project of a data warehouse system in a mobile phone company. Due to changing needs and requirements, it was necessary to readjust the previously deployed data warehouse. The main results of the project are the adherence of the system to the user's needs, the increasing use of the system and the increase of its trustworthiness.

Keywords: Data warehouse. Dimensional Modelling. Decision Support Sytems. Mobile Telephony. Information Systems.

LISTA DE FIGURAS

Figura 1 - Orientação por assunto	31
Figura 2 - Visão integrada do <i>data warehouse</i>	32
Figura 3 - Não-volatilidade do <i>data warehouse</i>	33
Figura 4 - Diferenças entre ODS e <i>Data warehouse</i>	36
Figura 5 - Esquema típico da Operational Data Store.....	37
Figura 6 - Arquitetura Inmon	40
Figura 7 - Evolução do sistema informacional segundo Inmon	41
Figura 8 - Exemplo de um ERD.....	42
Figura 9 - Nível intermediário (DIS).....	43
Figura 10 - Arquitetura Kimball.....	45
Figura 11 - Exemplo de aplicação da modelagem dimensional	48
Figura 12 - Exemplo de modelo estrela (<i>Star schema</i>)	49
Figura 13 - Exemplo de modelo floco de neve (<i>Snowflake</i>)	49
Figura 14 - Exemplo de aplicação da Matriz do <i>Data Warehouse Bus</i>	53
Figura 15 - Diferentes abordagens de <i>Data warehouse</i> - Inmon e Kimball.....	57
Figura 16 - Departamento financeiro da Empresa Beta	65
Figura 17 - Processo de negócio estudado	66
Figura 18 - Macro estrutura do ambiente do <i>data warehouse</i> da área financeira da Empresa Beta	70
Figura 19 - Exemplo de tela do Cognos Planning da Empresa Beta.....	74
Figura 20 – Exemplo de tela do Oracle BIEE	74
Figura 21 - Análise da arquitetura implantada	76
Figura 22 - Análise da escolha da melhor abordagem de <i>data warehouse</i>	77
Figura 23 - Análise ODS	79
Figura 24 - Análise dos componentes afetados pela readequação	81
Figura 25 - Fato de Custo Despesa antes do projeto	82
Figura 26 - Fato de Custo Despesa depois do projeto	83
Figura 27 - Fato de Custo Opex antes do projeto.....	83
Figura 28 - Fato de Custo Opex depois do projeto.....	84
Figura 29 - Fato de DRE antes do projeto.....	84
Figura 30 - Fato de DRE depois do projeto.....	85

LISTA DE TABELAS

Tabela 1 - Características Sistemas transacionais x Sistemas de suporte à decisão.....	29
Tabela 2 - Dados variantes no tempo	33
Tabela 3 - Principais diferenças entre as abordagens Inmon e Kimball.....	55
Tabela 4 - Principais critérios para a escolha da abordagem.....	59
Tabela 5 - Dimensões da qualidade de dados.....	62
Tabela 6 - Regionais da Empresa Beta.....	63
Tabela 7 - Classificações dos negócios da Empresa Beta	64
Tabela 8 - Análise das alterações necessárias na solução implantada.....	80
Tabela 9 - Dimensões da qualidade de dados aplicadas na Empresa Beta.....	87
Tabela 10 - Medidas propostas para resolver os problemas das dimensões da qualidade de dados	87

LISTA DE ABREVIATURAS E SIGLAS

ANATEL	Agência Nacional de Telecomunicações
DM	<i>Data mart</i>
DRE	Demonstração do Resultado do Exercício
DW	<i>Data Warehouse</i>
ERP	<i>Enterprise Resource Planning</i>
ETL	<i>Extract, Transform, and Load</i>
ODS	<i>Operational Data Store</i>
RH	Recursos Humanos
TI	Tecnologia da Informação

SUMÁRIO

1. INTRODUÇÃO.....	23
1.1 O estágio	23
1.2 A empresa	24
1.3 O problema	24
1.4 Objetivos e escopo	25
1.5 Estrutura geral.....	26
2. DATA WAREHOUSE	27
2.1 Perspectiva histórica	27
2.2 Definição.....	30
2.3 Objetivos do <i>Data Warehouse</i>	34
2.4 <i>Operational Data Store</i> (ODS).....	35
2.5 Metadados	37
2.6 Arquitetura - Inmon vs Kimball.....	38
2.6.1 Modelo Inmon.....	39
2.6.2 Modelo Kimball	44
2.6.2.1 Componentes de um <i>Data Warehouse</i> segundo Kimball	44
2.6.2.2 Modelagem de dados dimensional.....	47
2.7 Semelhanças e diferenças – Inmon vs. Kimball	54
2.7.1 Semelhanças.....	54
2.7.1.1 Dados com atributo de tempo	54
2.7.1.2 Processo de Extração, Transformação e Carga (Extract, Transform and Load – ETL)	54
2.7.2 Diferenças	55
2.7.2.1 Metodologias de desenvolvimento e arquiteturas.....	56
2.7.2.2 Diferenças na Modelagem de Dados	57
2.7.2.3 Diferenças filosóficas.....	58
2.8 Escolhendo a melhor abordagem	58
2.9 Qualidade dos dados do <i>Data warehouse</i>	60
3. ESTUDO DE CASO	63
3.1 A empresa	63
3.2 Área de Negócio	64

3.3 Visão geral do processo	66
3.4 O projeto	67
3.5 Problemas.....	69
3.6 Entendimento da situação anterior	69
3.7 Macroestrutura - Arquitetura	70
3.7.1 Sistemas Fontes.....	70
3.7.2 Área de <i>Staging</i>	72
3.7.3 Área de Apresentação dos dados	73
3.7.4 Ferramentas de acesso aos dados.....	73
3.8 Análises.....	75
3.8.1 Análises segundo a revisão bibliográfica.....	75
3.8.1.1 Análise da arquitetura	75
3.8.1.2 Análise da ODS	78
3.8.2 Análises do projeto	79
3.8.2.1 Modelagem dimensional.....	81
3.8.2.2 Qualidade dos dados	86
3.9 Conclusões	88
3.10 Resultados	89
3.11 Considerações Finais	89
4. CONCLUSÕES	91
5. REFERÊNCIAS BIBLIOGRÁFICAS	92

1. INTRODUÇÃO

Este capítulo visa apresentar as condições no qual o trabalho foi desenvolvido, descrevendo o estágio do autor, a empresa objeto de estudo, os problemas que o trabalho pretende solucionar, os objetivos, o escopo e a estrutura do trabalho.

1.1 O estágio

Este trabalho foi elaborado durante o período de estágio do autor em uma grande empresa multinacional do setor de tecnologia da informação, que terá seu nome ocultado devido ao pedido da mesma. Assim, a empresa será denominada como Empresa Alfa. O vínculo do estágio estendeu-se desde fevereiro de 2012 a novembro do mesmo ano.

A Empresa Alfa atua no mercado global, presente em mais de 170 países, oferecendo soluções de tecnologia de ponta a ponta, adequadas a empresas de todos os portes e perfis. A empresa opera em um modelo de empresa globalmente integrada e emprega mais de 430 mil pessoas no mundo. A subsidiária brasileira chegou ao país em 1917.

A empresa possui seu capital aberto, com ações negociadas na Bolsa de Valores de Nova York (NYSE). Sua subsidiária brasileira não negocia ações em bolsa. A empresa pode ser considerada uma multinacional de grande porte, com uma receita total global de U\$ 106,9 bilhões em 2011. No Brasil, conta com mais de 15 mil colaboradores, espalhados pelas principais cidades do país.

A Empresa Alfa atua em uma faixa ampla do mercado de Tecnologia da Informação, fornecendo tanto produtos como serviços. Entre as principais ofertas da organização, estão os serviços de terceirização (*outsourcing*) de qualquer aspecto da área de TI, consultoria de TI, consultoria de negócios, diversos softwares e aplicações de hardware.

O estágio foi realizado na área de consultoria de negócios da organização, mais especificamente na ênfase de otimização e analítica de negócios. Durante o período de estágio, o autor foi treinado em algumas ferramentas de modelagem financeira e obteve conhecimentos práticos do ambiente e dos componentes envolvidos em uma aplicação de *data warehouse*, atuando como suporte à equipe em diversas frentes.

1.2 A empresa

O presente trabalho foi desenvolvido em uma empresa do setor de telecomunicações móveis, com presença em todos os estados do país, atendendo a mais de 3600 municípios e com mais de 61 milhões de clientes. A operadora atende seus clientes com infra-estrutura e cobertura digital nas tecnologias 3G e GSM, e é controlada por um grupo estrangeiro, com acordos de roaming em mais de 160 países para serviços de voz e em mais de 140 países para tráfego de dados, nos cinco continentes.

A Empresa Beta, nome fictício utilizado no trabalho para preservar sua identidade, contratou os serviços da Empresa Alfa, na qual o autor realiza estágio, para atender às necessidades da área financeira da empresa. O presente trabalho irá focar nesse departamento específico da empresa, que será descrito com mais detalhes adiante. O autor do trabalho atuou como parte integrante de uma equipe de projeto na Empresa Beta.

1.3 O problema

A Empresa Beta contratou os serviços da Empresa Alfa no começo de 2010 para desenvolver um projeto que atendesse às necessidades de um setor específico da área financeira, responsável pelo orçamento e pelo fechamento gerencial mensal. A Empresa Alfa, na qual o autor realizou estágio, desenvolveu todo o ambiente e implementou a solução no mesmo ano, possibilitando à área realizar o planejamento e orçamento do ano de 2011, bem como apurar os resultados dos meses de 2010.

A solução envolvia uma ferramenta de modelagem financeira específica para a realização de orçamentos, bem como toda a infra-estrutura necessária para armazenar os dados gerados pelo processo, criando assim uma base com um grande histórico de informações, conhecida como *data warehouse*. O objetivo dessa base seria a de permitir consultas e geração de relatórios, que atuam como subsídios ao processo de tomada de decisões da organização, bem como a possibilidade de acompanhamento entre os valores realmente realizados contra os valores orçados.

Em uma indústria que está em constante transformação, como é o caso das empresas de telefonia móvel, alterações nas necessidades e nas regras de negócio podem e acontecem com o passar do tempo. O modelo inicialmente construído e implantado começou a mostrar limitações e problemas de aderência às necessidades da área de negócios.

Devido às mudanças internas em algumas diretrizes e regras de negócios, o modelo inicialmente implantado não conseguia mais atender aos interesses da área financeira, que passou a buscar alternativas que se adequariam melhor às suas necessidades, diminuindo a utilização do *data warehouse*. Outra questão levantada pelos usuários da área de negócios apontava para a falta de confiabilidade nos dados carregados na base, problema este que foi agravado com o passar do tempo e com a mudança de necessidades.

Com esse panorama de falta de aderência aos requisitos da área de negócios e crescente falta de confiança com relação aos dados carregados na base do *data warehouse*, a Empresa Beta voltou a contratar a Empresa Alfa, no início do ano de 2012, para realizar um projeto que iria readequar a solução para as novas necessidades internas e que contemplasse também a revisão da qualidade dos dados carregados.

1.4 Objetivos e escopo

O objetivo inicial do presente trabalho será o de apresentar a teoria referente ao tema de *data warehouse*, essencial para o entendimento da importância da utilização de uma base de dados única e integrada em uma organização e os benefícios esperados através dessa adoção. Como

objetivo final, o trabalho visa analisar o estudo de caso de um projeto de melhoria de um sistema de *data warehouse*.

O trabalho terá como foco estudar o caso de um setor específico da área financeira da Empresa Beta, não contemplando outros departamentos ou outros setores da área financeira. Será dado maior enfoque nas questões relativas ao ambiente e aos componentes do sistema de *data warehouse*, deixando de lado outros processos e sistemas da área.

1.5 Estrutura geral

O presente trabalho será dividido em 5 capítulos, como descritos em seguida.

O primeiro capítulo apresenta os objetivos do trabalho, as condições em que o mesmo foi desenvolvido, a empresa objeto de estudo, a definição do problema a ser tratado e a estrutura geral do trabalho de formatura.

O segundo capítulo apresenta uma revisão bibliográfica sobre *data warehouse*, estabelecendo os principais conceitos relativos ao tema do trabalho e apresentando as principais características e modelos que irão auxiliar o entendimento do estudo de caso apresentado no terceiro capítulo. É apresentada uma ampla pesquisa, em que foram consultados os principais autores do tema.

O terceiro capítulo apresenta o caso real de uma aplicação de *data warehouse* em uma empresa de telefonia móvel, na qual o autor deste trabalho trabalhou como consultor durante o período de estágio.

O quarto capítulo apresenta as principais conclusões obtidas durante o desenvolvimento do trabalho e do estudo de caso.

O quinto e último capítulo elenca todas as referências utilizadas para a elaboração deste trabalho, como livros, artigos, notas de aulas, entre outros materiais.

2. DATA WAREHOUSE

Este capítulo visa apresentar os principais conceitos relacionados ao tema do trabalho de formatura, primeiramente levantando uma perspectiva histórica sobre o conceito de *data warehouse*. Em seguida, o termo será definido e serão apresentados temas e características relevantes, bem como os principais modelos e abordagens sobre o tema.

2.1 Perspectiva histórica

O conceito de *data warehouse* remonta ao final da década de 1980, quando os pesquisadores da IBM Barry Devlin e Paul Murphy utilizaram o termo "business *data warehouse*" para designar a sua proposta de sistema de informações integrado. (HAYES, 2002).

Devlin e Murphy (1988) propõe em seu artigo para o IBM Systems Journal um modelo de sistema de informações que atue de forma consistente e integrada, com o objetivo de fornecer informações relevantes para os tomadores de decisão. Denominado "Business *data warehouse*", o sistema focaria no usuário final, que passaria a se concentrar mais em como utilizar a informações do que em como obter as informações.

Em 1990, W. H. Inmon publica seu primeiro livro abordando o tema, com o título de "Building the *Data Warehouse*". Por ser um dos primeiros livros a tratar do assunto, o autor obteve grande prestígio, sendo considerado por muitos o "pai" e criador do termo *data warehouse*. (BRESLIN, 2004). Porém, essa posição é questionada dentro do meio acadêmico, gerando discussões entre os diversos autores e correntes. (DEVLIN, 2009).

Percebe-se que a necessidade de um sistema que não somente registrasse os dados operacionais, mas que também os armazenasse e disponibilizasse para consultas, de forma organizada, motivou o início dos primeiros passos no desenvolvimento do conceito de *data warehouse*, sendo mencionado por diversos autores.

Machado (2006) afirma que a tecnologia de *Data Warehouse* é considerada como a evolução natural dos Sistemas de Apoio à Decisão (SAD). Sua crescente utilização pelas organizações foi motivada em grande parte pela necessidade de possuir informações estratégicas e como utilizá-las de modo a garantir respostas e ações rápidas, no ambiente atual altamente competitivo e mutável.

Inmon (1996) elenca os principais problemas relacionados à existência de diversas bases e sistemas de dados isolados dentro de uma organização:

- Falta de credibilidade dos dados

Cada sistema pode ter seus próprios dados, coletados em diferentes instantes de tempo e em diferentes pontos do sistema.

- Problemas de produtividade

Muitas fontes de dados e arquivos, que possuem diferentes *layouts* e configurações precisam ser analisadas para gerar um relatório, demandando um grande esforço para compilar todos esses dados.

- Incapacidade de transformar dados em informações

Os diversos sistemas isolados normalmente não possuem um grande histórico de dados, impossibilitando análises mais sofisticadas.

Segundo Kimball (2002), um dos ativos mais importantes de uma organização é a sua informação. Esse ativo normalmente é armazenado de duas maneiras em uma organização: em sistemas transacionais e no *data warehouse*. O autor propõe uma analogia para exemplificar esses diferentes papéis: os usuários do sistema transacional ou de operações da organização são aqueles que fazem a roda girar, enquanto que os usuários do *data warehouse* observam a roda girar.

Usuários do sistema transacional registram clientes, recebem pedidos, registram reclamações, por exemplo. Normalmente lidam com um registro de cada vez. Por outro lado, usuários do *data warehouse* contam os novos pedidos e comparam com os pedidos da semana passada e se perguntam por que conseguiram esses novos clientes e quais os motivos das reclamações

no período. Usuários do *data warehouse* costumam lidar com diversos registros ao mesmo tempo.

Deve ficar claro também a importância da divisão entre os dois sistemas, o transacional e o *data warehouse*, pois eles possuem necessidades e usuários distintos, que priorizam diferentes tipos de informações. Um *data warehouse* não pode ser uma mera cópia de um sistema transacional.

A Tabela 1 representa uma síntese das diferentes características entre os sistemas transacionais e os sistemas informacionais ou de suporte à decisão (*data warehouse*).

Sistemas transacionais	Sistemas de suporte à decisão (<i>Data warehouse</i>)
Orientado por aplicação	Orientado por assunto
Dados detalhados	Dados detalhados e sumarizados
Dados precisos, no momento do acesso	Representa valores no tempo, fotografias
Atende a comunidade funcional	Atende a comunidade gerencial
Atualizável	Não atualizável
Alta disponibilidade	Razoável disponibilidade
Estrutura estática	Estrutura flexível
Alta probabilidade de acesso	Média/baixa probabilidade de acesso
Suporte a operações do dia-a-dia	Suporte a necessidades gerenciais
Não há redundância	Redundância faz parte do sistema
Acesso a um único registro	Acesso a múltiplos registros de uma só vez
Dirigido à transação – OLTP	Dirigido à análise - OLAP
Pequena quantidade de dados utilizada em um processo	Alta quantidade de dados utilizada em um processo

Tabela 1 - Características Sistemas transacionais x Sistemas de suporte à decisão
Fonte: Inmon, 1996.

2.2 Definição

O *data warehouse* serve como um grande armazém de dados de uma organização, armazenando-os de forma organizada, a partir de diversas fontes de dados, sejam elas internas ou externas à organização. Permite que outros sistemas e ferramentas possam realizar consultas rápidas e precisas, fornecendo subsídios para o processo de tomada de decisões.

Kimball (2002, p. 397) define o termo *data warehouse* como sendo uma base de dados em que os dados operacionais são estruturados de tal forma que possam ser consultados e analisados facilmente.

Calvanese (2006), por sua vez, afirma que um *data warehouse* pode ser definido como sendo um conjunto de visões materializadas das fontes de informações operacionais de uma organização, desenhada para proporcionar apoio à análise de dados e às decisões gerenciais.

E Singh (2001) diz que um *data warehouse* é um conjunto de novos conceitos e ferramentas que evolui para uma tecnologia que permite atacar o problema de oferecer a todas as pessoas-chave da empresa acesso a qualquer nível de informação necessário para que a organização possa sobreviver e prosperar em um mundo cada vez mais competitivo.

Inmon (1996, p. 33), conhecido como o pai do *data warehouse*, propõe a definição clássica do termo da seguinte forma: “*Data Warehouse* é uma coleção de dados orientada por assunto, integrada, não-volátil e variante no tempo, que tem por objetivo suportar as decisões gerenciais.”.

A definição de Inmon é a mais popular e influenciadora para *data warehouses*. Apesar de não ser uma definição perfeita, ela enumera todas as funcionalidades essenciais que um *data warehouse* deveria fornecer a uma organização. (JIANG 2012)

De forma a clarificar a definição, o autor explica melhor cada um dos termos contidos na sua definição: (INMON, 1996; INMON e HACKATHORN, 1994)

- Orientada por assunto: é a orientação por assuntos ou por negócios da organização, retratando os principais assuntos de interesse da empresa, diferentemente dos sistemas transacionais tradicionais que são voltados para as aplicações com foco nas transações realizadas diariamente. Para os assuntos do *data warehouse*, deve-se focar nos dados que realmente sejam importantes para a tomada de decisão, análise e desempenho de processos ou atividades críticas.

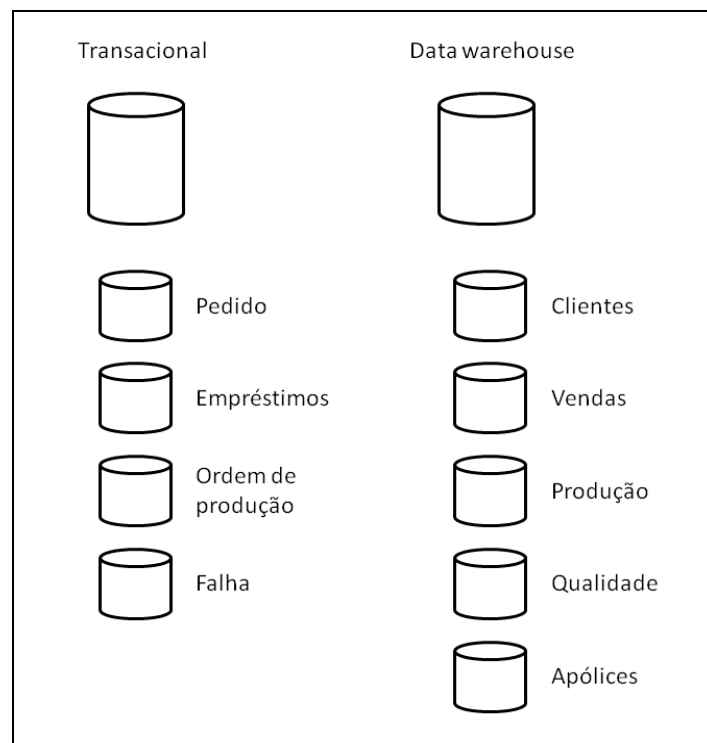


Figura 1 - Orientação por assunto
 Fonte: Autoria do próprio autor.
 Adaptado de Machado, 2006

- Integrada: é a característica mais importante de um *data warehouse*. A Figura 2 ilustra a integração que ocorre quando o dado passa do ambiente orientado por aplicações para o ambiente do *data warehouse*. Os dados são processados de modo a retirar inconsistências presentes entre as diversas aplicações, de modo a uniformizar e obter dados padronizados. A integração pode aparecer de diversas formas, como, por exemplo, na consistência das convenções de nomes, consistência na medição de variáveis, consistência nas estruturas de codificação, consistência de atributos físicos, consistência do formato das datas, entre outros.

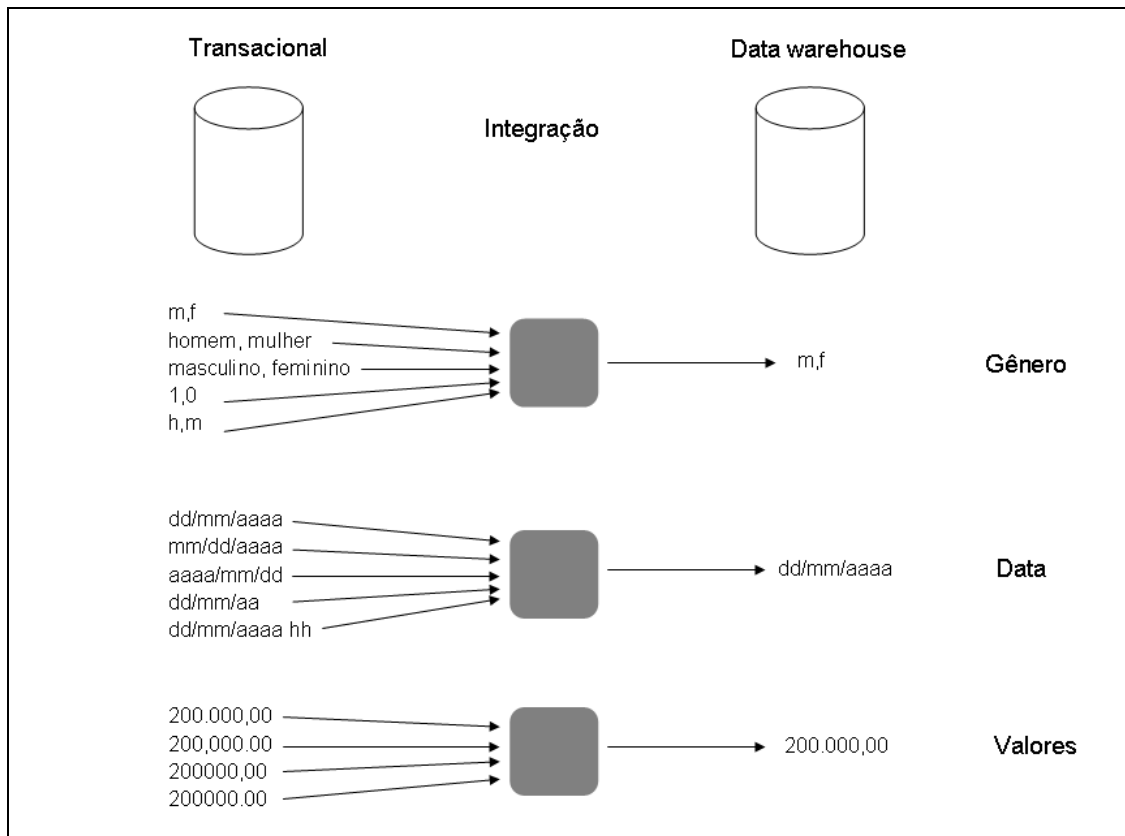


Figura 2 - Visão integrada do *data warehouse*

Fonte: autoria do próprio autor.

Adaptado de Kimball, 2002

- **Não-volátil:** os dados presentes no *data warehouse* somente permitem duas operações: consulta e carga (geralmente em massa), não ocorrendo nenhuma atualização dos dados no ambiente do *data warehouse*. Nos sistemas transacionais, a manipulação dos dados é constante, permitindo a inserção, a exclusão, a alteração e a consulta dos dados.

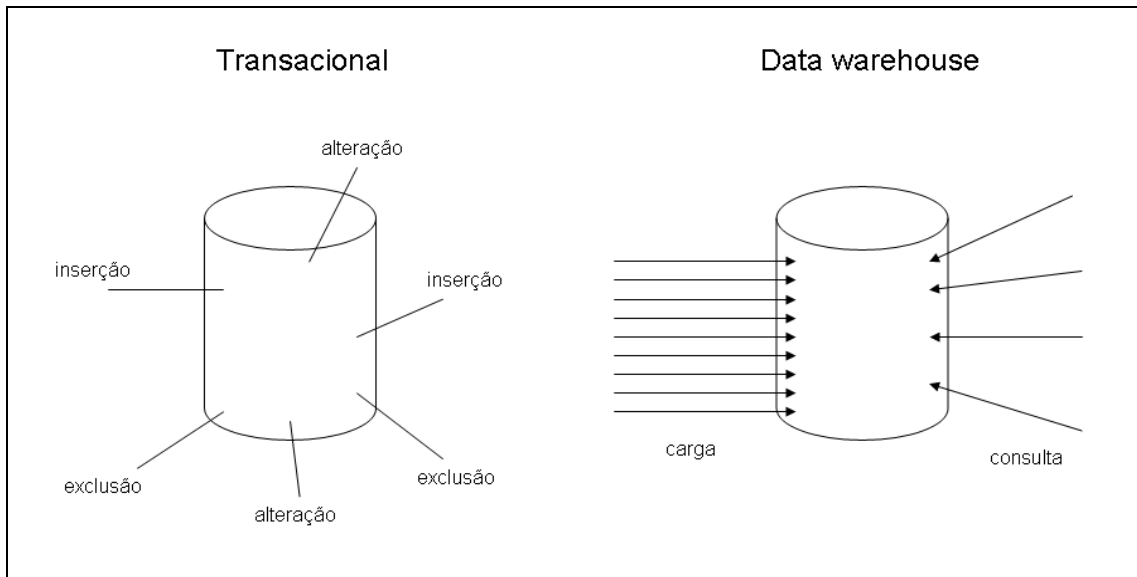


Figura 3 - Não-volatilidade do *data warehouse*
 Fonte: Inmon e Hackathorn, 1994

- Variante no tempo: os dados contidos no *data warehouse* são não atualizáveis e se referem a algum momento do tempo, funcionando como uma série de sofisticadas fotografias. No *data warehouse*, a estrutura de dados sempre contém algum elemento do tempo, e o horizonte de tempo costuma ser de 5 a 10 anos, em contraposição aos sistemas transacionais, que podem ou não conter elementos de tempo e tem horizontes de tempo entre 60 e 90 dias.

	Transacional	<i>Data warehouse</i>
Horizonte de tempo	60 a 90 dias	5 a 10 anos
Elemento de tempo	Presente ou não presente	Presente
Atualizável	Sim	Não

Tabela 2 - Dados variantes no tempo
 Fonte: Inmon, 1996.

Apesar da existência de diversas definições, opiniões e conceitos sobre o que seria *data warehouse*, Machado (2006) ressalta a importância de perceber que não existe um modelo de *data warehouse* pronto para ser utilizado. É necessário um trabalho de levantamento de necessidades da organização e de seus executivos, de modo a permitir a elaboração de um projeto de *data warehouse* que esteja alinhado com o direcionamento da empresa e de seus comandantes.

2.3 Objetivos do *Data Warehouse*

Segundo Kimball (2002), um *Data Warehouse*, ou Armazém de Dados, deve:

- **Permitir fácil acesso às informações da organização:** O conteúdo deve ser compreensível, isto é, as informações devem ser intuitivas e óbvias para o usuário da área de negócio.
- **Apresentar as informações da organização de forma consistente:** As informações dentro do armazém devem ter credibilidade. Não podem existir dois valores diferentes para a mesma informação.
- **Ser adaptável e flexível a mudanças:** Mudanças não podem ser simplesmente evitadas. As necessidades dos usuários, as condições de negócios, as informações e as tecnologias podem e normalmente se alteram com o passar do tempo. Mudanças no *data warehouse* não devem invalidar dados ou aplicações existentes.
- **Proteger as informações da organização:** Normalmente o *data warehouse* armazena dados sobre o que é vendido, para quem, e por qual preço. Tais informações são de extrema importância para uma organização, e poderiam causar grandes estragos nas mãos de pessoas erradas. Desse modo, o *data warehouse* deve ter um controle efetivo sobre o acesso às suas valiosas informações.
- **Servir como ponto inicial para melhores tomadas de decisão:** Deve servir como um verdadeiro sistema de suporte a decisão (SSD, ou em inglês DSS (*Decision Support System*)), auxiliando tomadas de decisão que tenham real impacto no negócio.
- **Obter aceitação dos usuários das áreas de negócio da organização:** De nada adianta utilizar os melhores produtos e plataformas se a área de negócio não adotar o *data warehouse* e utilizá-lo por pelo menos seis meses. Normalmente a utilização do *data warehouse* é opcional, diferentemente do que costuma ocorrer com a implantação de um novo sistema de operações, fato este que pode contribuir para a não aceitação do *data warehouse*.

Para obter sucesso na implementação de um *data warehouse*, é necessário ter um pé na área da Tecnologia da Informação (TI) e outro na área de negócio que utilizará o *data warehouse*. A aceitação pela área de negócio é medida por quanto do *data warehouse* é utilizado, o que está intimamente ligado com a facilidade de uso, ou em inglês, *user-friendly*.

2.4 Operational Data Store (ODS)

Outro conceito importante quando lidamos com sistemas informacionais e *data warehouses* é a *Operational Data Store* (ODS). A ODS é um sistema híbrido entre os sistemas transacionais e o *data warehouse*, apresentando características de ambos os sistemas. A ODS é considerada uma extensão do *data warehouse* para o mundo dos sistemas transacionais, e deve ficar claro que ela pertence ao ambiente de operações da organização.

Inmon e Hackathorn (1994) definem ODS como sendo uma coleção de dados orientada por assuntos, integrada, volátil e atual ou quase atual, que tem por objetivo suportar decisões operacionais do dia a dia.

Comparando com a definição de *Data warehouse* apresentada no tópico 2.2, pode-se perceber que elas diferem nos dois últimos pontos da definição. A ODS é volátil, ao contrário do *data warehouse*, pois seus dados podem ser atualizados, e o *data warehouse* possui um longo histórico de dados, variante no tempo, ao contrário da ODS que possui somente dados atuais ou quase atuais.

Outro ponto de diferenciação entre os dois sistemas, não muito evidente à primeira vista, está na quantidade de dados armazenados, com o *data warehouse* contendo muito mais informações que a ODS. O público alvo dos dois sistemas também difere, sendo a ODS voltada para usuários mais operacionais e funcionais, ao passo que o *data warehouse* visa atender as necessidades do público gerencial de uma organização. As principais diferenças elencadas entre a *Operational Data Store* e o *Data Warehouse* estão representadas na Figura 4.

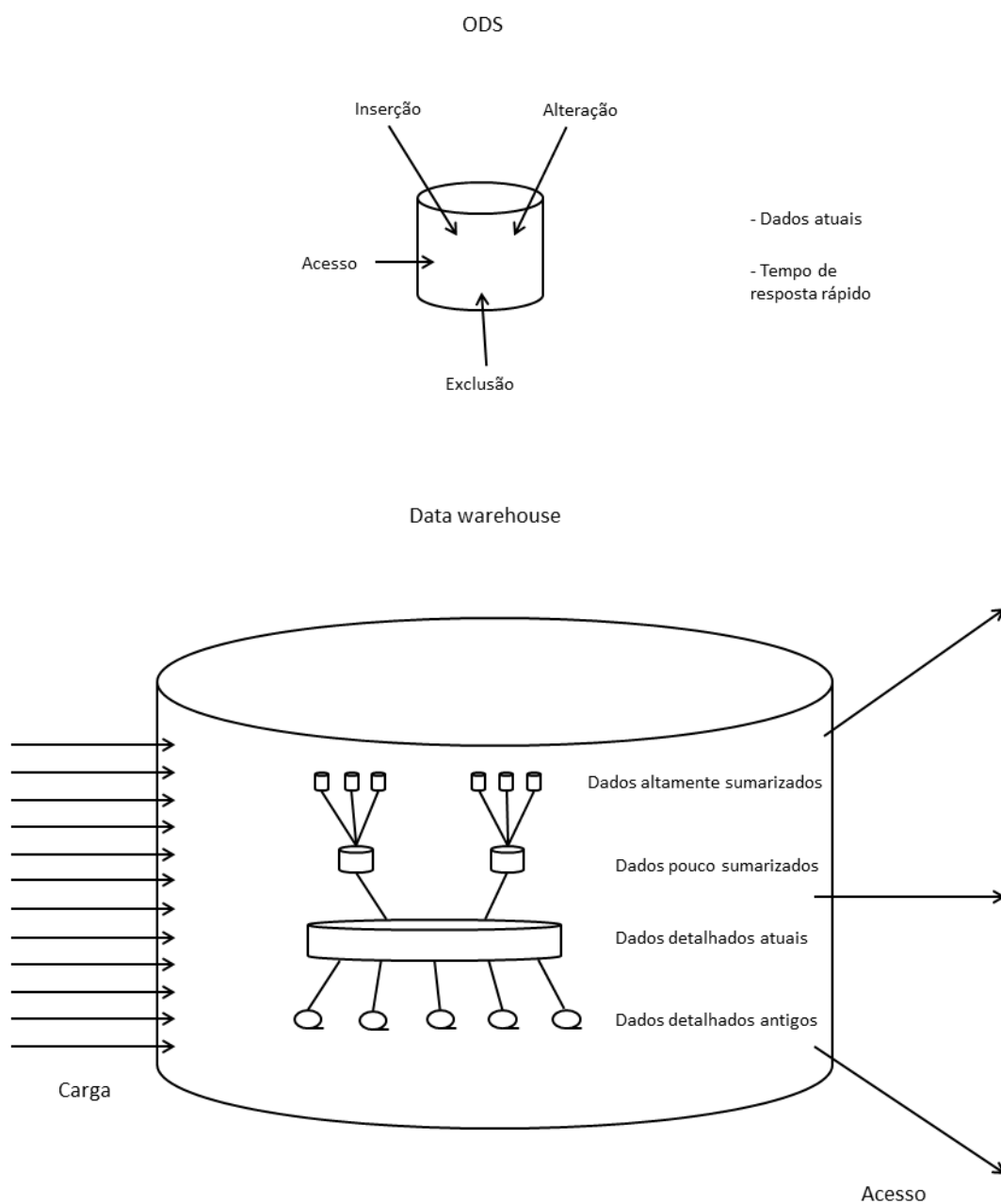


Figura 4 - Diferenças entre ODS e *Data warehouse*
 Fonte: Autoria do próprio autor.
 Adaptado de Inmon e Hackathorn, 1994.

Em geral, as organizações recorrem à adoção da ODS, pois ela exige um menor esforço no processo de implementação e proporciona um alívio quase que imediato para aquelas com dificuldades de integração de dados em seu ambiente transacional.

A ODS serve como uma agregadora de dados dos diversos sistemas “legados” transacionais que as organizações desenvolveram ao longo de sua existência, normalmente atendendo às

necessidades imediatas e pecando pela falta de integração entre os diversos sistemas. Desse modo, a ODS funciona como uma etapa anterior à carga dos dados no *Data Warehouse*, transformando os dados recebidos dos diversos sistemas. Segundo Inmon e Hackathorn (1994), deve ficar clara a separação entre a ODS e o *data warehouse*, pois em hipótese nenhuma os dois podem ser combinados.

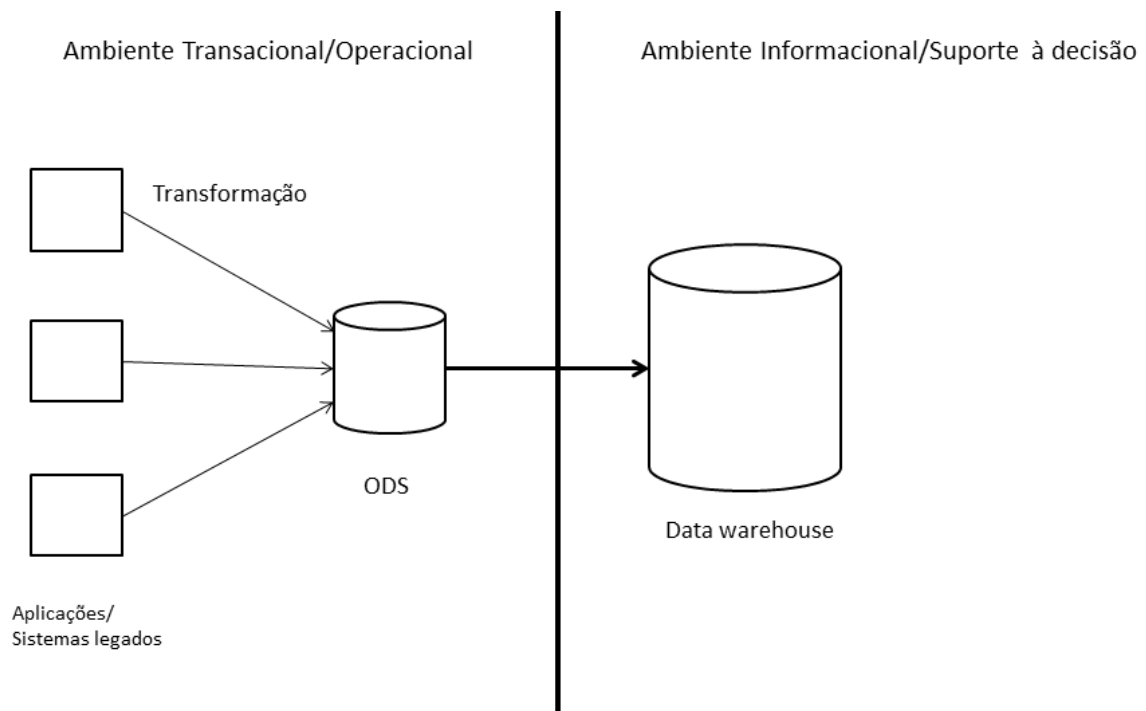


Figura 5 - Esquema típico da Operational Data Store
Fonte: Inmon e Hackathorn, 1994.

Durante o processo de transformação, as seguintes atividades podem ser executadas: conversão de dados, decisão de qual a melhor fonte dos dados, sumarização dos dados, decodificação/codificação, alteração das estruturas dos dados, formatação, recálculo, entre outras.

2.5 Metadados

A clássica definição de que metadado é simplesmente "dados sobre os dados" é muito simplificada, afirma Harris (2011). Para Kimball (2002), "metadado é toda a informação no ambiente do *data warehouse* que não é o próprio dado". O autor ainda destaca a importância

de considerar os metadados em um *data warehouse*, apesar de sua estruturação exigir grandes esforços, pois a mesma será de grande ajuda na administração do sistema.

O metadado pode ser entendido como uma etiqueta que fornece a definição, a descrição e o contexto para o dado. Apesar de ser muitas vezes negligenciada, existe um forte relacionamento entre o metadado e a qualidade do dado, e por extensão, entre o metadado e a tomada de decisão baseada nos dados ¹(HARRIS, 2011).

De acordo com Kimball (2002) e Harris (2011), os metadados, entre outras aplicações, envolvem:

1. Informações sobre os dados dos sistemas de operações ou transacionais.
2. Documentação de todo o processo de ETL (*Extraction, Transformation and Load*), como, por exemplo, as regras de limpeza e transformação, rotinas de execução, *layout* das tabelas.
3. Documentação dos componentes do *data warehouse*, dos *data marts*, das tabelas, das dimensões. Inclui aspectos como permissões de segurança, configurações, estatísticas de uso, documentação do usuário.
4. Definições das tabelas relacionais.
5. Modelos de dados conceituais e lógicos.

2.6 Arquitetura - Inmon vs Kimball

Muitas organizações hoje em dia necessitam criar *data warehouses* para suportar o processo de tomada de decisão com dados históricos. Essas organizações enfrentam uma série de diferentes opções, tanto em termos de ferramentas de software como em relação às alternativas de desenvolvimento. Possuir um bom conhecimento sobre os dois principais modelos de arquitetura de *data warehouse* é essencial para realizar as melhores escolhas para a organização, observando quais características organizacionais se enquadram melhor em cada abordagem, de forma a permitir que a organização desenvolva um modelo apropriado a

¹ Em inglês, Data-driven decision making.

suas necessidades e a suas características. As duas diferentes abordagens mais conhecidas e utilizadas para *data warehouse* são a do Inmon e a do Kimball.

Em geral, *data warehouses* contém *terabytes* de dados que devem estar disponíveis para consulta pelos usuários finais. A grande quantidade de dados é explicada pela necessidade de um grande histórico de dados, tanto sumarizados quanto detalhados. Os dados em sua maioria tem origem nos sistemas transacionais internos da organização, e após passar por um processo de transformação, são armazenado nas tabelas relacionais do *data warehouse*. (KIMBALL, 2002)

Uma famosa anedota sobre a utilização dos dados contidos em um *data warehouse* é aquela em que uma cadeia de varejo americana descobriu que pais de primeira viagem normalmente levam cerveja e fraldas na mesma compra, uma relação que não seria normalmente esperada à primeira vista. A tecnologia de *data warehouse* é creditada como responsável por essa descoberta. (ALBERT, 2000)

2.6.1 Modelo Inmon

Segundo Inmon (1996), o ambiente informacional de uma organização pode ser dividido em quatro níveis: operacional/transacional, *data warehouse*, departamental (*data mart*) e individual. Os três últimos níveis constituem o *data warehouse*. A Figura 6 representa o modelo proposto pelo autor.

Arquitetura Inmon – Ambiente arquitetado

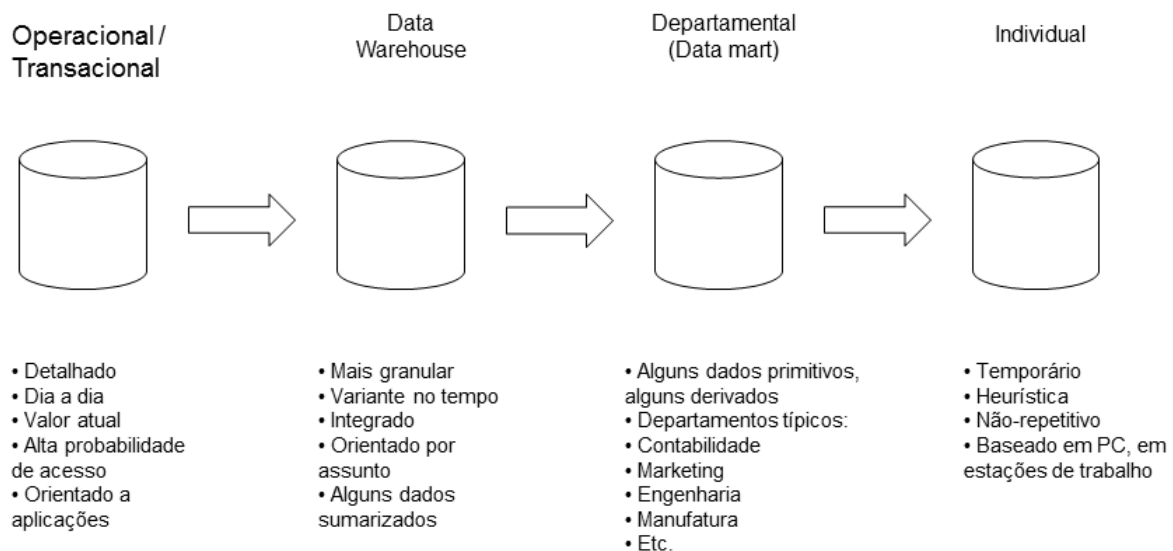


Figura 6 - Arquitetura Inmon
Fonte: Inmon, 1996.

Para Inmon (1996), os dados presentes nos diversos sistemas transacionais de uma organização passam por transformações e manipulações de modo a poderem ser movidos para o nível do *data warehouse*. Esse processo é conhecido como *Extract, Transform and Load* (ETL), e será abordado com mais detalhes no tópico 2.7.1.2. A passagem dos dados do nível de *data warehouse* para o departamental envolve a sumarização dos dados, em diferentes níveis de agrupamento, dependendo das necessidades do departamento específico. Apesar dos diferentes níveis de sumarização, a arquitetura proposta por Inmon garante a consistência dos dados entre os diversos departamentos, pois todos os dados tem origem na mesma fonte, o *data warehouse*.

O nível individual é criado pelos usuários finais do ambiente arquitetado, que criam conjuntos de dados para suprir o processo de tomada de decisões. Em geral, esse quarto e último nível tende a ser temporário e armazenado nos computadores pessoais (PCs) dos usuários.

Inmon (1996, p. 43-45) defende uma abordagem de implementação gradual de um *data warehouse*, no qual no começo operam somente os sistemas operacionais, com uma crescente utilização do ambiente de *data warehouse*. Com o passar do tempo, começam a aparecer os primeiros ambientes departamentais, denominados de *data marts*. Após um determinado

período de tempo, a arquitetura está amplamente desenvolvida. O processo é exemplificado na Figura 7.

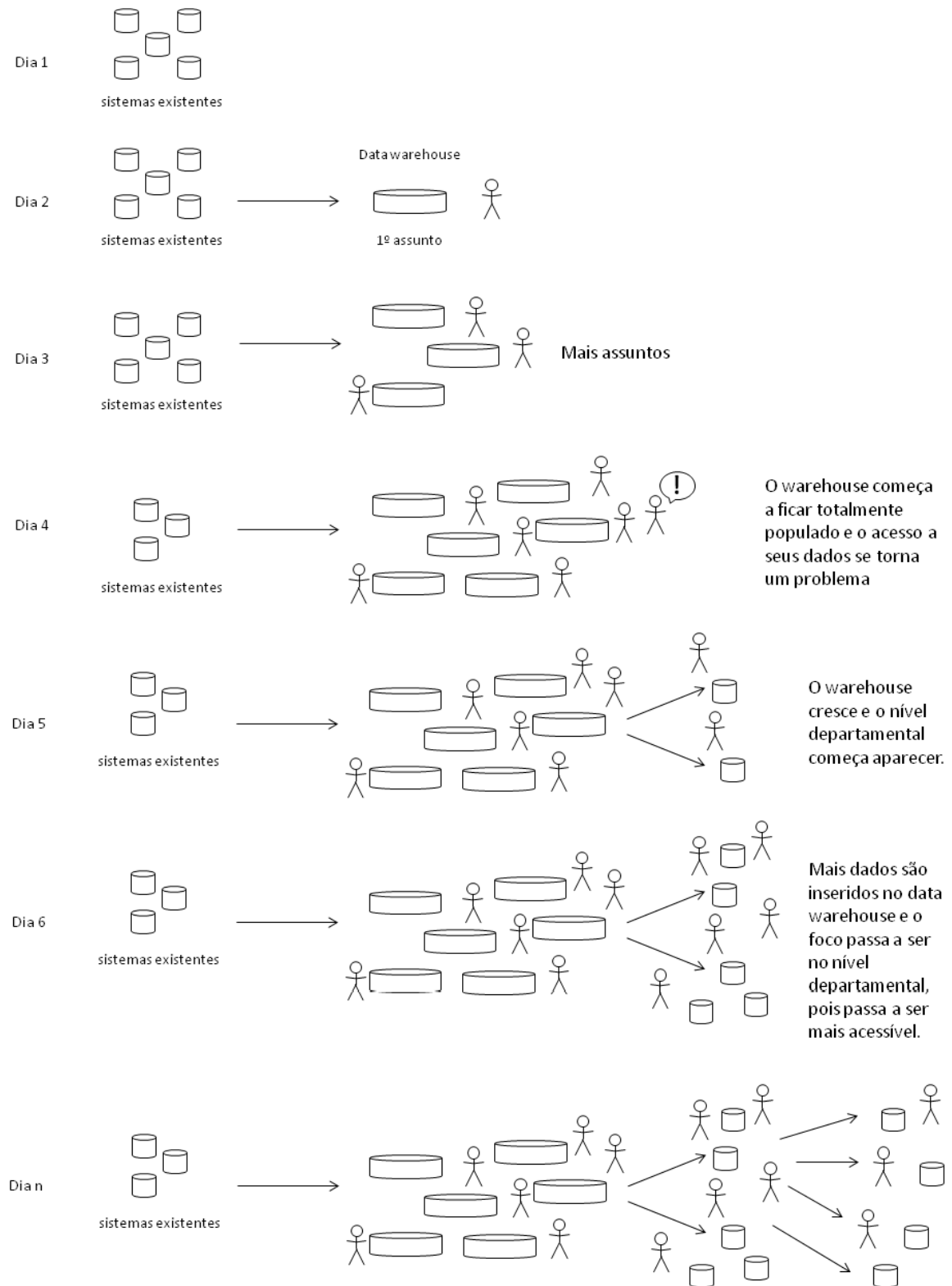
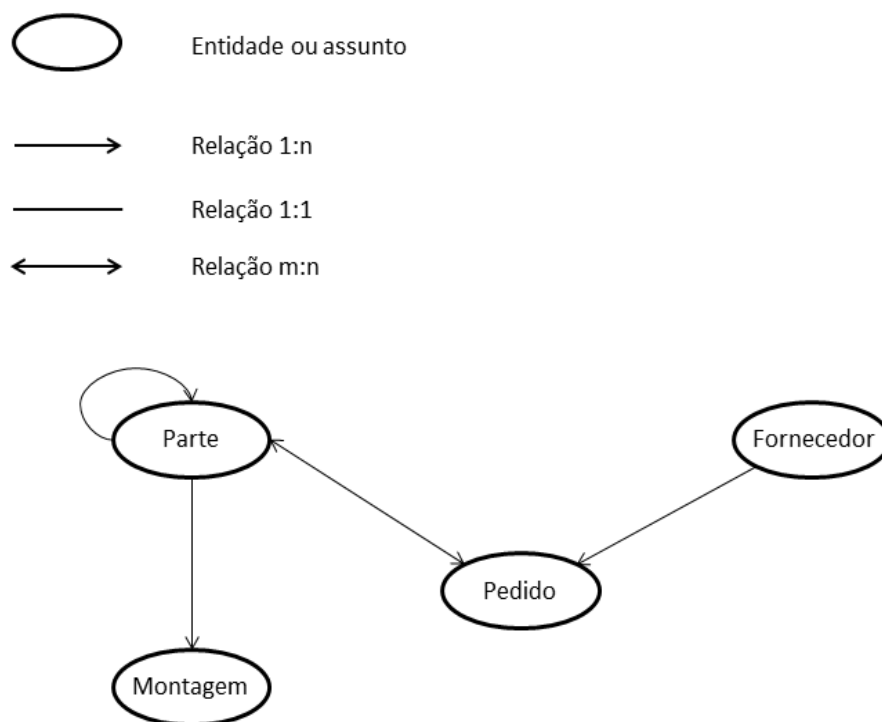


Figura 7 - Evolução do sistema informacional segundo Inmon
Fonte: Inmon (1996).

Inmon (1996) defende que o esforço para implementar o *data warehouse* vale a pena pois possibilita a criação de diversas bases de dados departamentais a partir da mesma base, eliminando os riscos de incompatibilidade entre os dados dos diferentes departamentos.

O modelo de dados proposto por Inmon é composto por três níveis, sendo eles: alto nível ou ERD, nível intermediário ou DIS (data item set) e o nível baixo ou modelo físico. O primeiro nível é composto por *Entity Relationship Diagrams* (ERDs) ou Diagramas de relacionamentos de entidades, que apresentam o maior nível de abstração do modelo. Os ERDs são utilizados para denotar quais são as entidades ou os assuntos envolvidos no modelo de dados a ser considerado e qual o relacionamento entre eles. São utilizados no desenvolvimento de bases de dados operacionais também. A soma dos diversos ERDs de cada departamento forma o ERD da organização.



Exemplo de um ERD para o ambiente de manufatura

Figura 8 - Exemplo de um ERD
Fonte: Inmon, 1996.

O segundo nível do modelo estabelece um *Data item set* (DIS) ou conjunto de artigos de dados para cada entidade ou assunto do primeiro nível, a fim de detalhar cada entidade. Existem quatro componentes no nível intermediário: grupo de dados primário, grupo de dados

secundário, conector (representando o relacionamento entre os dados e o assunto do ERD), e tipo de dado. Dentro de um modelo DIS, cada retângulo representa uma tabela lógica, e as conexões entre as tabelas são as mesmas que conectam entidades nos ERDs. Retângulos abaixo do grupo primário representam um grupo secundário de dados. O tipo de dado é representado por outras colunas de retângulos à direita. Um exemplo de uma instituição financeira é apresentado na Figura 9 de modo a exemplificar cada componente do nível intermediário.

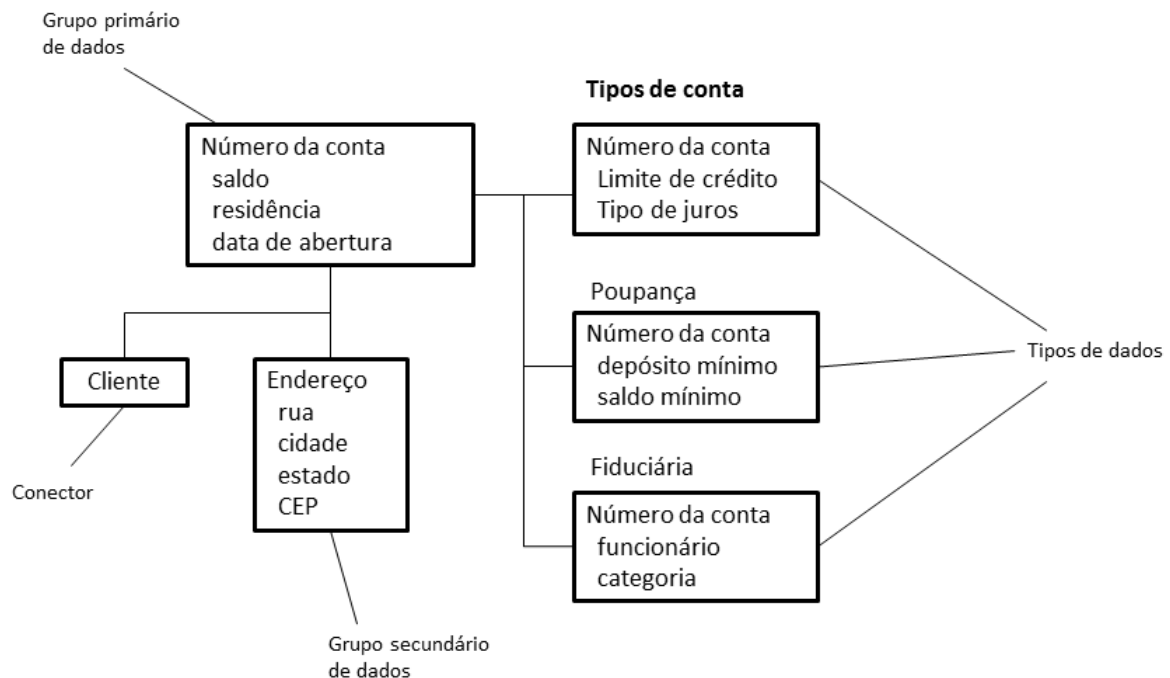


Figura 9 - Nível intermediário (DIS)

Fonte: Inmon, 1996.

O último nível do modelo de dados de Inmon é o físico. Ele é criado a partir do nível intermediário, simplesmente estendendo o modelo DIS para incluir chaves e características físicas ao modelo, de modo a se parecer com uma série de tabelas, conhecidas como tabelas relacionais. O modelo físico está muito próximo de como deverão ser construídas as tabelas, ficando pendente apenas a consideração de questões relativas ao desempenho do sistema.

Segundo Breslin (2004), a arquitetura de Inmon pode ser considerada como sendo evolucionária, e não revolucionária, pois Inmon acredita que tanto o *data warehouse* quanto as bases transacionais fazem parte de um conjunto maior, que inclui todo o sistema de

informações de uma organização. Desse modo, o *data warehouse* para Inmon se apoia em muitas práticas, princípios e padrões dos sistemas transacionais, sendo assim mais uma evolução do que revolução. Além do mais, o público primário de Inmon são os profissionais de Tecnologia da Informação, pois é exigido certo conhecimento tecnológico para utilizar as ferramentas propostas. Assim, os usuários finais do *data warehouse* teriam um papel predominantemente passivo no desenvolvimento do DW.

2.6.2 Modelo Kimball

Segundo Breslin (2004), o modelo proposto por Kimball difere em alguns pontos importantes da abordagem tradicional, proposta por Inmon. Um das diferenças significativas está na utilização de uma metodologia única para o modelamento dos dados, conhecida como modelagem dimensional. Outro ponto de diferenciação está na arquitetura geral de Kimball, que prevê múltiplas bases de dados que devem ser altamente interoperáveis, através da utilização do Data Bus e de dimensões coerentes e conformes. Cada um dos tópicos de diferenciação (Modelagem dimensional e Data Bus e dimensões coerentes) será abordado com mais detalhes em tópicos seguintes.

2.6.2.1 Componentes de um *Data Warehouse* segundo Kimball

Para Kimball (2002), existem quatro componentes distintos que são considerados ao se explorar o ambiente de um *Data Warehouse*: Sistemas operacionais, Área de *Staging*, Área de Apresentação dos Dados e Ferramentas de Acesso aos Dados. É importante compreender as partes que compõem um *Data Warehouse* antes de começar a combiná-las. Cada componente serve para uma função específica. Uma das maiores ameaças para o sucesso de um *Data Warehouse* é a possibilidade de confundir os papéis e as funções dos componentes. A Figura 10 apresenta uma representação do modelo proposto pelo autor.

Arquitetura Kimball

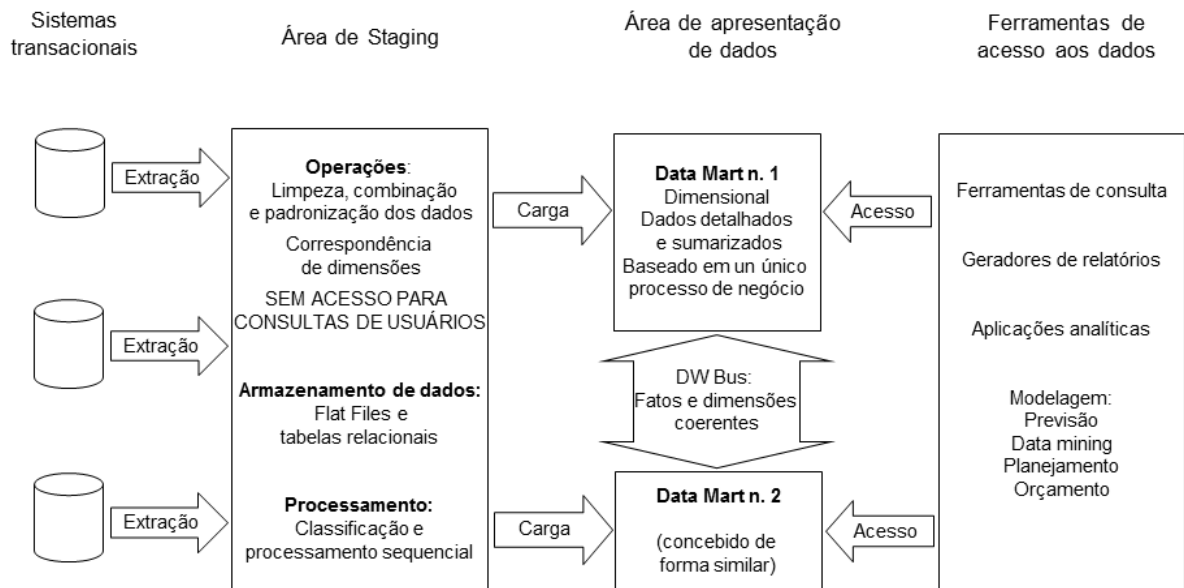


Figura 10 - Arquitetura Kimball
Fonte: Kimball, 2002.

Sistemas transacionais

São os sistemas operacionais responsáveis por capturar e armazenar as transações do negócio. Esses sistemas possuem pouco histórico de dados, o que pode ser aliviado com a implantação de um bom *data warehouse*. Cada sistema operacional normalmente atua de modo isolado, compartilhando pouca ou nenhuma informação com sistemas de outras áreas da organização.

Área de Staging

A Área de Staging é uma combinação de uma área de armazenamento com uma série de processos normalmente conhecidos por Extração-Transformação-Carga (ETL, em inglês, *Extract-Transformation-Load*). A Área de Staging é tudo o que está presente entre os Sistemas Transacionais e a Área de Apresentação dos Dados. Tal área não deve ser acessível

para os usuários das áreas de negócios. O primeiro passo desse processo é a Extração dos dados das fontes de dados transacionais, isto é, a leitura e o entendimento dos dados e a cópia dos dados necessários ao *data warehouse*. Em seguida, os dados extraídos podem ser transformados de diversas maneiras, como por exemplo limpeza dos dados (correção de erros ortográficos, tratar elementos ausentes, formatação para um modelo padrão da organização), combinar dados de diversas fontes, retirar dados duplicados e atribuir chaves de warehouse. Ao final, os dados são carregados no ambiente do *data warehouse*, em um *data mart*.

Área de Apresentação dos Dados

É a área na qual os dados são organizados, armazenados e disponibilizados para consulta. Para os usuários da área de negócios, a Área de Apresentação dos Dados é o *data Warehouse*, pois eles só entram em contato com esta parte da base de dados. Normalmente, em um modelo Kimball, a Área de Apresentação dos Dados é composta por uma série de *data marts* integrados. Um *data mart* representa somente uma parte da Área de Apresentação dos Dados, sendo a representação de um único processo de negócio.

Considerações sobre a Área de Apresentação dos Dados

Kimball (2002) defende que os dados devem ser armazenados em Modelos Dimensionais, ao invés dos tradicionais modelos normalizados, também conhecidos como 3NF. Os modelos normalizados são extremamente úteis para o desempenho do processamento transacional, pois exigem a alteração ou inserção em somente um ponto do modelo. Porém, para o propósito de facilidade de consulta que um *data warehouse* deve possuir, as tabelas normalizadas são muito complicadas e complexas para os usuários finais. Desse modo, a modelagem dimensional permite entendimento por parte do usuário final, desempenho e resiliência para mudanças. A modelagem dimensional será abordada em seguida, no tópico 2.6.2.2.

Outro ponto defendido pelo autor diz respeito aos dados contidos em um *data mart*, que devem ser detalhados e atômicos. Os dados devem possuir granularidade compatível com as necessidades dos usuários da área de negócio, que são imprevisíveis e estão sempre mudando.

E o último ponto defendido por Kimball fala sobre a importância de que todos os *data marts* sejam construídos utilizando dimensões e fatos comuns, pois esse é o ponto chave da Arquitetura Data Bus desenvolvida pelo autor. Com *data marts* construídos com dimensões comuns a todos, é possível integrá-los formando um modelo de *data warehouse* descentralizado.

Diferentemente do que defende a corrente tradicionalista sobre *data warehousing*, os *data marts* modernos podem ser atualizados, em alguns casos até frequentemente. Obviamente, dados incorretos devem ser corrigidos, mas também podem ocorrer mudanças de classificação, hierarquia e status. (KIMBALL, 2002)

Ferramentas de Acesso aos Dados

Trata-se de um termo amplo, que engloba uma variedade de recursos que podem ser oferecidos aos usuários da área de negócios para a tomada de decisão analítica. Por definição, todas as ferramentas de acesso aos dados fazem consultas na área de apresentação dos dados do *data warehouse*. Dentre a ampla gama de opções, as ferramentas podem variar desde uma simples ferramenta de consulta até complexos sistemas de modelagem e previsão.

2.6.2.2 Modelagem de dados dimensional

A modelagem dimensional pode parecer estranha para profissionais de TI, que normalmente estão acostumados à modelagem relacional. A modelagem dimensional começa com tabelas, que podem ser tanto tabelas fato ou tabelas de dimensão. As tabelas fato contêm métricas e valores, enquanto as tabelas dimensionais contêm atributos das métricas carregadas nas tabelas fato. A modelagem dimensional viola as regras de normalização impostas pela modelagem relacional, para atingir um alto nível de desempenho enquanto mantém o ambiente acessível aos usuários finais.

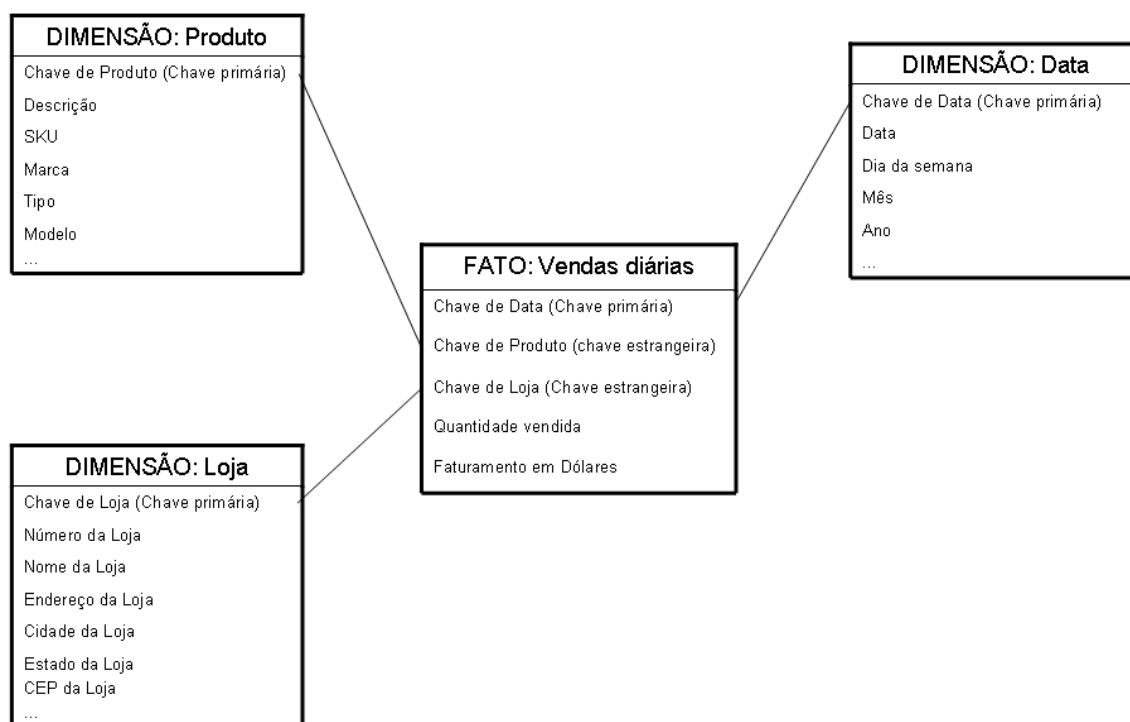


Figura 11 - Exemplo de aplicação da modelagem dimensional
Fonte: Kimball (2002)

A partir do exemplo demonstrado na Figura 11, pode-se perceber a simplicidade e a simetria proporcionadas pela modelagem dimensional. Os usuários finais das áreas de negócio se beneficiam da simplicidade ao tornar os dados mais fáceis de serem entendidos e de se orientar, quando comparados a modelos mais tradicionais normalmente utilizados em desenvolvimento de bases de dados. Além de proporcionar facilidade para os usuários, a simplicidade também impacta no desempenho, otimizando o processo e a velocidade de resposta.

O modelo apresentado na Figura 11 é um típico exemplo de um modelo estrela ou *Star schema*, em que a fato se encontra no centro da estrela e as dimensões relacionadas a essa fato ficam arranjadas ao redor da entidade central. Outro exemplo do modelo estrela é apresentado na Figura 12.

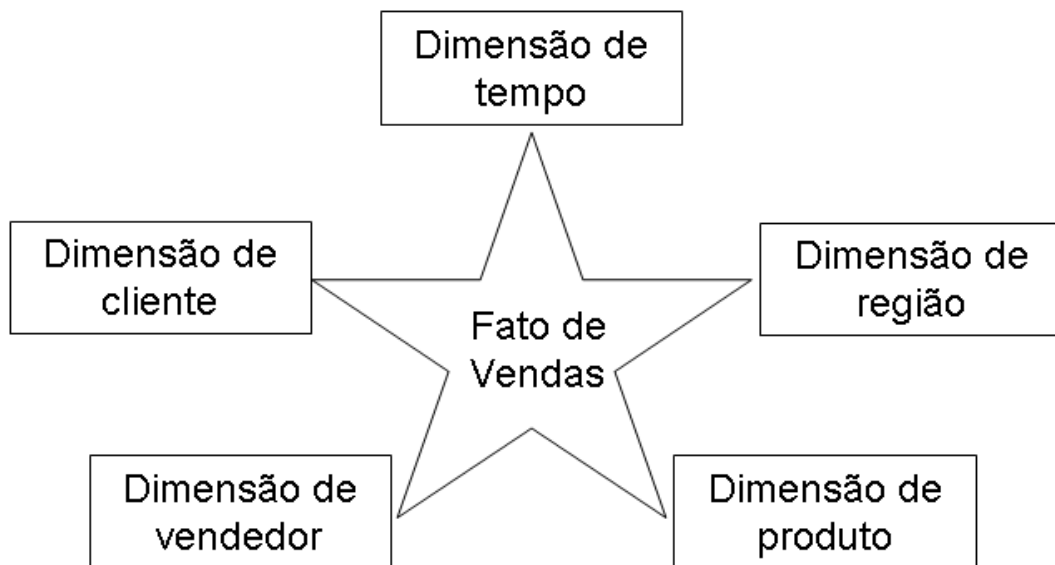


Figura 12 - Exemplo de modelo estrela (*Star schema*)
 Fonte: Machado (2006).

Outro modelo conhecido para representar modelos multidimensionais é o modelo Floco de neve ou *Snowflake*. Nesse modelo, uma ou mais dimensões são decompostas em suas hierarquias. Na realidade, trata-se de uma tentativa de aproximação com os modelos normalizados ou 3NF, que evitam a redundância entre tabelas. Um exemplo do modelo floco de neve é apresentado na Figura 13.

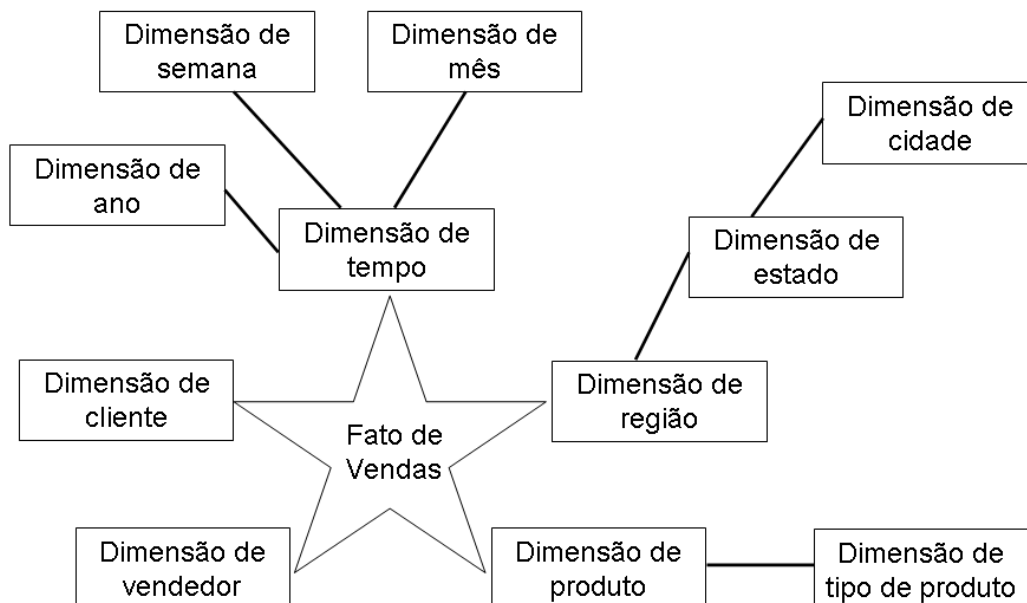


Figura 13 - Exemplo de modelo floco de neve (*Snowflake*)
 Fonte: Machado (2006)

Ainda com relação à modelagem dimensional dos dados, é importante notar a necessidade de manter tabelas fato com poucas colunas, pois normalmente tais tabelas apresentam muitas linhas de dados. Ao acrescentar uma nova coluna, aumenta-se muito o espaço ocupado pela tabela. Por outro lado, as tabelas das dimensões apresentam poucas linhas quando comparadas às fatos. Assim, em geral as dimensões possuem muitas colunas, pois representam os diversos atributos que as métricas podem assumir.

O alto nível de desempenho é alcançado ao transferir para as tabelas dimensionais os filtros e as restrições que os usuários possam necessitar, para então realizar uma única consulta à tabela fato. Como os filtros ficam nas tabelas dimensionais, que são bem menores que as tabelas fato, o usuário apresenta um ganho de desempenho enorme ao chegar às tabelas fato já munido de todas as condições que o satisfaçam.

2.6.2.3 O processo de modelagem dimensional de quatro etapas

A abordagem de Kimball envolve uma abordagem *bottom-up*, isto é, a construção de um *data mart* de cada vez, que ao serem somados, se transformam no *data warehouse*.

Kimball (2002) define um processo de modelagem dimensional de quatro etapas, sendo as etapas:

- Selecionar o processo de negócio
- Definir o grão
- Escolher as dimensões
- Identificar as fatos

A definição de processo de negócio proposta por Kimball é bastante abrangente. Kimball (2002) define processo de negócio como sendo uma atividade de negócio desempenhada em uma organização que é tipicamente suportada por um sistema de coleta de dados. Para tornar clara a definição de processo de negócio, o autor apresenta alguns exemplos: vendas no ponto

de venda, estoque, pedidos, entregas, etc. É importante não confundir processo de negócio com departamentos organizacionais ou funções organizacionais. O autor recomenda que o primeiro processo a ser escolhido para o *data warehouse* seja aquele que possua maior impacto e que consiga responder às necessidades mais urgentes da organização.

Em seguida, deve-se definir o nível de granularidade requerido, isto é, o nível de detalhe requerido para os dados presentes no *data warehouse*. O menor nível de granularidade é chamado de atômico, significando que os dados não podem ser mais subdivididos. Para Kimbal (2002), deve-se preferencialmente desenvolver modelos dimensionais para a informação mais atômica possível, para um dado processo de negócio. O *data warehouse* normalmente demanda que os dados estejam expressos no menor nível possível de grão para cada dimensão.

O próximo passo no processo proposto por Kimball é o de escolha das dimensões. Uma pergunta que exemplifica a escolha de dimensões é: “Como os usuários da área de negócios descrevem seus dados resultantes do processo de negócio?”. É preciso identificar um conjunto robusto de dimensões que representem todas as possíveis descrições que um valor pode ter em uma tabela fato. Após a escolha de cada dimensão, deve-se listar todos os atributos que podem aparecer em uma tabela de dimensão. Exemplos comuns de dimensões incluem data, produto, consumidor, tipo de transação e status.

O quarto e último passo é determinar quais fatos serão incluídos nas tabelas fato. Fatos podem ser determinadas ao se responder a seguinte pergunta: “O que estamos medindo?”. Os usuários finais da área de negócios estão bastante interessados em analisar o desempenho das medidas de desempenho do processo de negócio. Fatos típicas normalmente envolvem números que podem ser somados, como por exemplo quantidade pedida, custos, despesas, receita, etc.

2.6.2.3 Data Warehouse Bus e dimensões coerentes e conformes

No modelo de Kimball, os dados são copiados dos sistemas transacionais para a Área de *Staging*. Na Área de *Staging*, os dados são transformados de modo a ficarem consistentes e

compatíveis com as consultas dos usuários finais. Da Área de *Staging*, os dados são carregados nos *data marts*. E os usuários finais consultam os dados nos *data marts*.

Cada *data mart* é baseado em um único processo de negócios. Alguns exemplos de processos de negócio são ponto de venda, estoque, gestão de pedidos, etc. Pode acontecer que mais de um departamento tenha participação em um determinado processo de negócios, resultando em que não existe um único dono de um certo *data mart*.

O *Data Warehouse Bus* é a parte da arquitetura de Kimball que permite que a soma dos diversos *data marts* formem uma base verdadeiramente integrada, isto é, o *data warehouse*. A Arquitetura *Data Warehouse Bus* é outra forma de dizer que todos os *data marts* precisam utilizar as mesmas dimensões padronizadas e conformes. Para serem consideradas conformes, as dimensões precisam assegurar que as chaves, os nomes das colunas, as definições dos atributos e os valores dos atributos sejam consistentes através de todos os processos de negócios. Kimball (2002) define que duas dimensões são conformes quando elas são exatamente as mesmas, ou então quando uma é um subconjunto da outra.

Uma ferramenta que auxilia no processo de criação, documentação e comunicação da arquitetura *Data Warehouse Bus* é a “Matriz do *Data Warehouse Bus*”, que é ilustrada na Figura 14.

Processos de negócio	Dimensões comuns							
	Data	Produto	Loja	Promoção	Armazém	Fornecedor	Contrato	Transportadora
Vendas de varejo	x	x	x	x				
Estoque do varejo	x	x	x					
Entregas do varejo	x	x	x					
Estoque do armazém	x	x			x	x		
Entregas do armazém	x	x			x	x		
Ordens de compra	x	x			x	x	x	x

Figura 14 - Exemplo de aplicação da Matriz do *Data Warehouse Bus*
 Fonte: Kimball, 2002.

As linhas da matriz representam *data marts* da organização, cada um representando um processo de negócios. Todos os *data marts* identificados devem ser incluídos na matriz, sejam eles *data marts* primários ou *data marts* consolidados. As colunas da matriz representam as dimensões comuns através da empresa. É interessante elencar e criar o maior número possível de dimensões antes de preencher a matriz, pois isso facilita o processo de determinar se certa dimensão poderia ser associada a um certo *data mart*. As células marcadas indicam que a dimensão da coluna está relacionada ao processo de negócio da linha.

Existem duas maneiras possíveis de ler a Matriz do *Data Warehouse Bus*. Ao percorrer as linhas, é possível obter uma visão clara e rápida de quais dimensões estão associadas a cada *data mart*. A outra visão é mais útil ainda, ao denotar interação entre os *data marts* e as dimensões comuns.

A matriz serve como uma poderosa ferramenta tanto para o planejamento quanto para a comunicação. Ao analisar a matriz, é possível perceber quais dimensões devem ter prioridade e devem receber maiores atenções para a questão da conformidade e coerência. É eficaz também na comunicação entre as diversas áreas de uma organização, tanto horizontalmente quanto verticalmente.

2.7 Semelhanças e diferenças – Inmon vs. Kimball

As duas principais semelhanças entre os modelos (a existência de dados com atributo de tempo e processo de ETL) são necessárias para a existência dos sistemas de suporte à decisão. Desse modo, Breslin (2004) afirma que os dois modelos são similares somente nas características em que tinham que ser similares, e que se diferenciam profundamente nas demais áreas.

2.7.1 Semelhanças

Primeiramente serão tratados os aspectos nos quais os dois modelos são semelhantes e que possibilitam a existência dos sistemas de *data warehouse*.

2.7.1.1 Dados com atributo de tempo

O atributo de tempo é a característica que mais define os dados de um *data warehouse*. Ambas as arquiteturas ressaltam a importância do fator tempo nos dados do *data warehouse*, que em geral devem ser armazenados por 5 a 10 anos. Kimball define o atributo de tempo como sendo a “dimensão de data”. Por outro lado, Inmon define como o “elemento de tempo”. Qualquer que seja a escolha de abordagem, os usuários finais podem realizar consultas por dia, mês, trimestre, ano, feriados, dias de semana, fins de semana, etc.

2.7.1.2 Processo de Extração, Transformação e Carga (Extract, Transform and Load – ETL)

O ambiente de *data warehouse* começa com o processo de ETL. Os dados são extraídos das bases de dados transacionais, transformadas para atender aos padrões do *data warehouse*, e

por fim são carregados, tanto na base única do *data warehouse* de Inmon ou então nas diversas pequenas bases de dados, conhecidas como *data marts*, propostas por Kimball.

O processo de ETL é essencial para a viabilidade do *data warehouse*, pois este é o responsável por garantir a integridade dos dados dentro do DW. Os sistemas transacionais e legados não são desenhados para produzirem resultados compatíveis entre si, o que torna a tarefa de integrar os dados no *data warehouse* extremamente trabalhoso, sendo algumas vezes considerada mais intensiva do que as próprias atividades de análise de suporte à decisão.

2.7.2 Diferenças

As diferenças entre os modelos de Inmon e Kimball são muitas e de grande importância. As principais diferenças estão nas áreas de metodologias de desenvolvimento, modelagem de dados e arquiteturas de *data warehouse*. A Tabela 3 apresenta um resumo das diferenças entre as duas abordagens. Em seguida, apresenta-se uma breve descrição dos principais pontos.

	Inmon	Kimball
Metodologia e arquitetura		
Abordagem geral	Top-down	Bottom-up
Complexidade do método	Bastante complexo	Relativamente simples
Discussão do desenho físico	Complexo	Superficial
Modelagem dos dados		
Orientação dos dados	Orientação por assuntos	Orientação por processos
Ferramentas	Tradicionais, baseadas nas bases de dados operacionais	Modelagem dimensional
Acessibilidade do usuário final	Pequena	Alta
Filosofia		
Público	Profissionais de TI	Usuários finais
Objetivo	Entregar uma solução técnica baseada em comprovados métodos e tecnologias de banco de dados	Entregar uma solução que facilite o acesso do usuário final à informação, em um tempo resposta razoável

Tabela 3 - Principais diferenças entre as abordagens Inmon e Kimball
Fonte: Breslin, 2004.

2.7.2.1 Metodologias de desenvolvimento e arquiteturas

O modelo de Inmon de *data warehouse* atômico implica na utilização de algum grau de desenvolvimento *top-down*, que costuma envolver considerável grau de complexidade. Como o *data warehouse* atômico deve atender a organização inteira, todas as bases de dados departamentais obtêm seus dados através do *data warehouse* atômico.

De modo geral, a metodologia e a arquitetura de Inmon tem uma orientação mais técnica, tendo como principal interesse assegurar que a solução técnica funcione. Poucos usuários das áreas de negócios possuem a bagagem necessária para entender a abordagem de desenvolvimento de Inmon, deixando o controle e domínio por parte da equipe de TI.

Por outro lado, a metodologia de modelagem dimensional de quatro etapas de Kimball é bastante acessível para os usuários finais. Os conceitos de *Data warehouse Bus* e dimensões conformes podem ser entendidos pelos usuários finais sem grandes esforços de estudo. Por definição, uma abordagem *bottom-up* como a do Kimball envolve menos elementos que a *top-down*. O escopo mais restrito do *data mart* é consideravelmente mais fácil para os usuários compreenderem.

A Figura 15 exemplifica as diferenças de abordagens de arquitetura entre os dois modelos. No modelo de Inmon, os dados são extraídos dos sistemas transacionais, passam pelo processo de ETL, sendo transformados e carregados na base de dados única do *data warehouse*. Em seguida, caso seja do interesse da organização, os dados contidos no *data warehouse* central são passados para as bases de dados departamentais, os *data marts*. Pode-se perceber que a metodologia de Inmon pressupõe um desenvolvimento centralizado, fato esse que implica em uma coerência maior entre os dados das bases departamentais, pois elas possuem a mesma fonte.

A abordagem de Kimball é mais direta e mais rápida de ser implementada, como pode ser observado na figura. As diversas bases de dados departamentais, os *data marts*, são carregados diretamente pelos sistemas transacionais, após os dados passarem pelo processo de ETL. Para conferir um caráter de *data warehouse* e de base de dados única e integrada de uma organização, Kimball propõe a utilização do *Data warehouse Bus* e das dimensões coerentes,

de modo a minimizar problemas de duplicidade de informação proveniente da arquitetura descentralizada de Kimball. Cabe ressaltar que a união dos diversos *data marts* formam o *data warehouse*, não existindo um *data warehouse* fisicamente único e integrado.

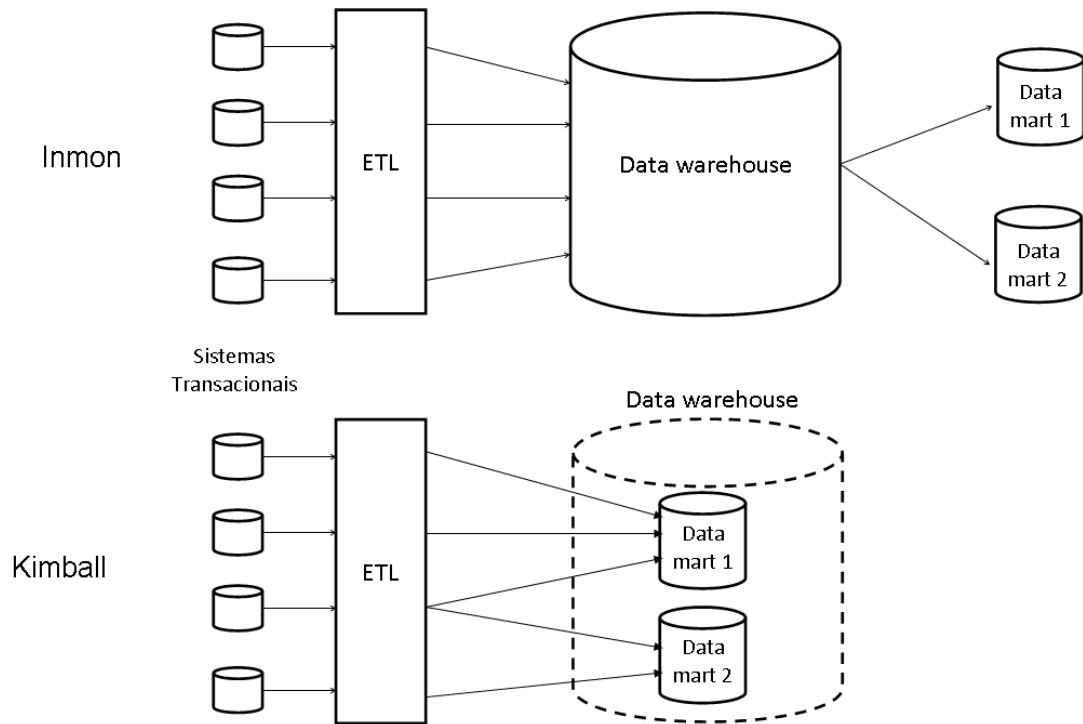


Figura 15 - Diferentes abordagens de *Data warehouse* - Inmon e Kimball
Fonte: Autoria do próprio autor.

2.7.2.2 Diferenças na Modelagem de Dados

Dois pontos com claras diferenças entre a modelagem de dados para Inmon e para Kimball são a orientação dos dados e as técnicas e regras de modelagem.

Com relação à orientação dos dados, Inmon adota uma abordagem orientada aos assuntos ou então também conhecida como direcionada pelos dados, isto é, a natureza dos dados direciona o processo de modelagem de dados. Para desenvolver a modelagem de dados, são utilizadas ferramentas tradicionais, que em geral necessitam de um treinamento especial que os usuários

finais não possuem, ficando a cargo da área de TI a modelagem dos dados em um modelo Inmon.

Por outro lado, a abordagem de Kimball utiliza uma orientação por processos, isto é, a modelagem é definida através do estudo dos dados de um processo de negócios, que em geral atravessa as fronteiras departamentais de uma organização. Essa abordagem é condizente com a nova abordagem proposta por Kimball, da modelagem dimensional, em que o processo determina quais serão as métricas (fatos) e os atributos (dimensões).

2.7.2.3 Diferenças filosóficas

Como já deve ter ficado claro agora, Inmon considera a área de TI como a principal desenvolvedora e mantenedora do *data warehouse*. O autor acredita que o desempenho máximo será atingido ao assegurar que o processo de desenvolvimento tenha um perfil mais técnico. Por outro lado, Kimball acredita que os usuários finais e os profissionais de TI compartilham papéis iguais no processo de desenvolvimento, e que essa participação ativa dos usuários finais na solução aumenta a chance deles aceitarem o *data warehouse* quando completo.

2.8 Escolhendo a melhor abordagem

Breslin (2004) propõe a utilização de oito critérios de avaliação para auxiliar no processo de determinação de qual abordagem será utilizada no desenvolvimento de um *data warehouse*. A Tabela 4 apresenta tais critérios e quais características se enquadram melhor em um modelo ou em outro.

Característica	Favorece Kimball	Favorece Inmon
Natureza das decisões	Táticas	Estratégicas
Requisitos de integração de dados	Áreas de negócios individuais	Integração em toda a empresa
Estrutura dos dados	Métricas de negócio, medidas de desempenho, <i>scorecards</i>	Dados que não são métricas; dados que devem atender diferentes e variadas necessidades
Escalabilidade	Necessidade de adaptação a requisitos altamente voláteis, dentro de um escopo limitado	Escopo crescente e requisitos mutantes são fatores críticos
Persistência dos dados	Sistemas fonte são relativamente estáveis	Sistemas fonte possuem alta taxa de mudanças
Requisitos de habilidades e pessoal	Pequenas equipes de generalistas	Grande equipe(s) de especialistas
Tempo para entrega	Necessidade da primeira aplicação de <i>data warehouse</i> é urgente	Requisitos da organização permitem um tempo maior até a primeira aplicação
Custo de implantação	Baixo custo inicial, com projetos subsequentes custando aproximadamente o mesmo	Alto custo inicial, porém com menores custos para projetos subsequentes

Tabela 4 - Principais critérios para a escolha da abordagem

Fonte: Breslin, 2004.

De modo geral, uma organização irá obter sucesso ao utilizar a abordagem de Inmon se possuir uma grande equipe de especialistas em *data warehouse*, planeja um grande projeto para atender as necessidades de dados de toda a empresa, armazena dados que não são métricas de negócio e pode esperar um tempo maior até observar os primeiros resultados – em geral, de quatro a nove meses. Essas características e requisitos se alinham com a recomendação de Inmon de primeiro construir uma infra-estrutura sólida e abrangente, que envolva a organização como um todo.

Por outro lado, a abordagem de Kimball é mais apropriada para organizações com características diferentes. Se existir urgência para a primeira aplicação de *data warehouse*, a

abordagem Kimball é a mais indicada, pois ela é considerada mais ágil que a de Inmon, quando considerado o tempo de entrega do primeiro *data mart* contra a primeira base de dados departamental do modelo de Inmon. O modelo de Kimball também é indicado caso a organização tenha maior facilidade em alocar pequenas equipes de generalistas para o desenvolvimento do *data warehouse* e possua a intenção de armazenar métricas de negócio.

É importante ressaltar que a escolha entre uma ou outra abordagem não se resume à simplificação feita acima, porém esta deve servir de base para a discussão das necessidades e características do *data warehouse* de determinada empresa.

Por fim, pesquisas têm demonstrado que possuir o conjunto de *soft skills* certo pode ser tão importante quanto as habilidades técnicas e o conhecimento. Eckerson (2003) afirma que projetos não são bem sucedidos por causa de seu desenho inovador ou por uma nova tecnologia radical, mas sim devido às habilidades sociais ou *soft skills*, como por exemplo liderança, comunicação, planejamento e relacionamentos interpessoais.

Eckerson (2003) continua afirmando que independente da abordagem a ser escolhida, é essencial que a equipe envolvida com o desenvolvimento do *data warehouse* utilize as habilidades sociais deliberadamente e de forma eficaz. Isso envolve assegurar que a organização possua uma visão bem articulada sobre o papel e o uso do *data warehouse*, e que aloque recursos suficientes para criar e manter o *data warehouse*. Essas responsabilidades não costumam ser da equipe de desenvolvimento de TI, mas são críticas para o sucesso do projeto do *data warehouse*.

2.9 Qualidade dos dados do *Data warehouse*

A qualidade dos dados em uma organização tem se tornado cada vez mais um fator crítico, com o recente movimento de aumento da disponibilidade de dados e informações. Para um ambiente de *data warehouse*, o tema se torna ainda mais importante, pois o *data warehouse* suporta a tomada de decisões táticas ou estratégicas, que possuem grande impacto nas organizações. Uma decisão errada tomada com base em informações incorretas pode provocar sérias consequências, como perdas financeiras e inclusive a retirada da empresa do mercado.

Um problema que a má qualidade de dados provoca, muito pior que o custo direto, é o descrédito interno e externo e suas consequências sobre os processos de tomada de decisão e sobre a percepção que os clientes e fornecedores formam acerca dos sistemas da empresa. Por outro lado, a qualidade dos dados da organização pode ser considerada uma vantagem competitiva. Quando isso acontece, é possível identificar rapidamente potenciais oportunidades de negócio ou marketing a partir de transformações e análises sobre os dados comerciais atuais e históricos. Ainda são poucas as organizações que adotam práticas estruturadas de gestão de dados e, sem esta prática, é certa a exposição a riscos e falhas nas iniciativas de Tecnologia de Informação e de negócio. (MATTIODA, FAVARETTO, 2006).

Segundo Orr (1998), existem seis regras gerais para a qualidade de dados em sistemas informacionais:

1. Dados não utilizados não podem permanecer corretos por muito tempo;
2. A qualidade dos dados em um sistema de informações é função de seu uso, e não de sua coleção;
3. A qualidade dos dados não será melhor que o seu uso mais rigoroso;
4. Problemas de qualidade dos dados tendem a piorarem à medida que o sistema envelhece;
5. Quanto menor a probabilidade de um atributo de um dado mudar, mais traumático será quando ele finalmente mudar;
6. As leis da qualidade de dados se aplicam tanto aos dados quanto aos metadados.

Almeida Jr (2011) elenca as principais causas da baixa qualidade em sistemas informacionais:

- Entrada de dados inicial com problema;
- Degradação dos dados, tornando-os obsoletos;
- Utilização incorreta dos dados: não conhecimento do significado do dado ou sua má interpretação;
- Alta incidência de mudanças e reestruturação: troca de base de dados, migração de ambientes, etc.
- Baixo nível de monitoramento dos dados.

Ainda segundo Almeida Jr (2011), a qualidade dos dados pode ser dividida em algumas dimensões que a representam: acuracidade (correção, livre de erros), disponibilidade, atualização e confiabilidade/credibilidade. Essas dimensões auxiliam a entender e a avaliar a qualidade dos dados em uma determinada aplicação. A Tabela 5 apresenta a definição de cada uma das dimensões.

Dimensão	Definição
Acuracidade	Quanto a informação é correta.
Disponibilidade	Quanto a informação está disponível, ou fácil e rapidamente recuperável.
Atualização	Quanto a informação está suficientemente atualizada para a tarefa a ser realizada.
Confiabilidade/Credibilidade	Quanto a informação é considerada como verdadeira e verossímil.

Tabela 5 - Dimensões da qualidade de dados
Fonte: Adaptado de Mattioda (2009).

De modo a combater a falta de qualidade dos dados em um *data warehouse*, Almeida Jr (2011) afirma que o nível de qualidade dos dados pode ser melhorado pelo incentivo ao uso dos dados por parte dos usuários da área de negócios. O autor apresenta uma analogia com o fenômeno biológico da atrofia, em que a pouca utilização ou total inutilização de ossos e músculos provoca a sua redução, até a eventual morte dos tecidos.

Orr (1998) corrobora com essa postura, afirmando que a única maneira de melhorar a qualidade dos dados é melhorando o seu uso. Deve-se focar na identificação exata de como os dados serão utilizados, a fim de elencar os usuários e envolvê-los no processo como um todo, de modo a criar uma sensação de responsabilidade sobre os dados por parte dos usuários. Dessa maneira, os usuários se empenham mais em manter a qualidade dos dados alta.

Outro aspecto relevante levantado por Orr (1998) está na importância da educação e do treinamento para manter a alta qualidade de dados no longo prazo. Em geral, é difícil fazer os usuários e os gerentes entenderem os fundamentos da qualidade de dados. Entretanto, assim que começam a perceber os efeitos positivos do uso consciente dos dados, se tornam rapidamente adeptos das melhores práticas e do uso eficaz dos dados.

3. ESTUDO DE CASO

Nesse capítulo será apresentado o estudo de caso real realizado em uma empresa do setor de telecomunicações, mais especificamente de telefonia móvel. Primeiramente será descrito o ambiente no qual o estudo se encaixa, bem como os problemas associados a esse cenário. Em seguida, é feita uma abordagem do caso levando em conta toda a revisão bibliográfica desenvolvida no capítulo 2. No final, são apresentadas as conclusões relativas ao caso estudado.

3.1 A empresa

Como descrito no item 2.1, a Empresa Beta atua no setor de telefonia móvel, em todos os estados brasileiros. Ela é controlada por um grupo estrangeiro, para a qual toda a operação brasileira se reporta.

A organização divide as suas operações pelo país em algumas regionais, que estão representadas na Tabela 6.

Regional	
Matriz	Matriz
SP	Regional São Paulo
RJ/ES	Regional Rio de Janeiro
NE	Regional Nordeste (exceto BA e SE)
CO	Regional Centro Oeste
RS	Regional Rio Grande do Sul
PR/SC	Regional Paraná/Santa Catarina
BA/SE	Regional Bahia/Sergipe
MG	Regional Minas Gerais
NO	Regional Norte

Tabela 6 - Regionais da Empresa Beta
Fonte: Documento interno.

Outra classificação relevante de como a Empresa Beta divide suas operações é evidenciada pelos componentes de Mercado e de Plataforma/Produto, apresentados na Tabela 7.

Mercado	
Consumo	
Corporativo	
Plataforma/Produto	
Pós-Pago	Conta
	Controle
	Banda Larga Pós
	Rastreamento e Telemetria
Pré-Pago	Pré-Pago Sem Banda Larga Pré
	Banda Larga Pré

Tabela 7 - Classificações dos negócios da Empresa Beta
Fonte: Documento interno.

3.2 Área de Negócio

A área de negócio da Empresa Beta em que o trabalho terá enfoque é o departamento financeiro da organização, mais especificamente a área conhecida como Informações Financeiras, responsável pelo fechamento gerencial mensal, pelo planejamento e orçamento financeiro de toda a empresa e por análises financeiras. Outra atribuição da área é a geração de relatórios contábeis gerenciais, que atuam como suporte ao processo de tomada de decisão.

A área de fechamento mensal é responsável por consolidar os dados obtidos da contabilidade e tratá-los de modo a representar os resultados de cada divisão, de cada linha de produto e de cada mercado, por exemplo. Desse modo, a área é responsável por traduzir os números contábeis em métricas apropriadas que permitem tomar decisões mais acertadas.

Outra parte da equipe da área é responsável pelo planejamento e pelo orçamento de toda a organização, recebendo os dados das diversas regionais e diretorias para compor as previsões de gastos do ano corrente e do próximo ano. As análises financeiras são realizadas tanto por membros da equipe do fechamento quanto por membros do orçamento.

A Figura 16 apresenta um retrato do departamento financeiro da Empresa Beta.

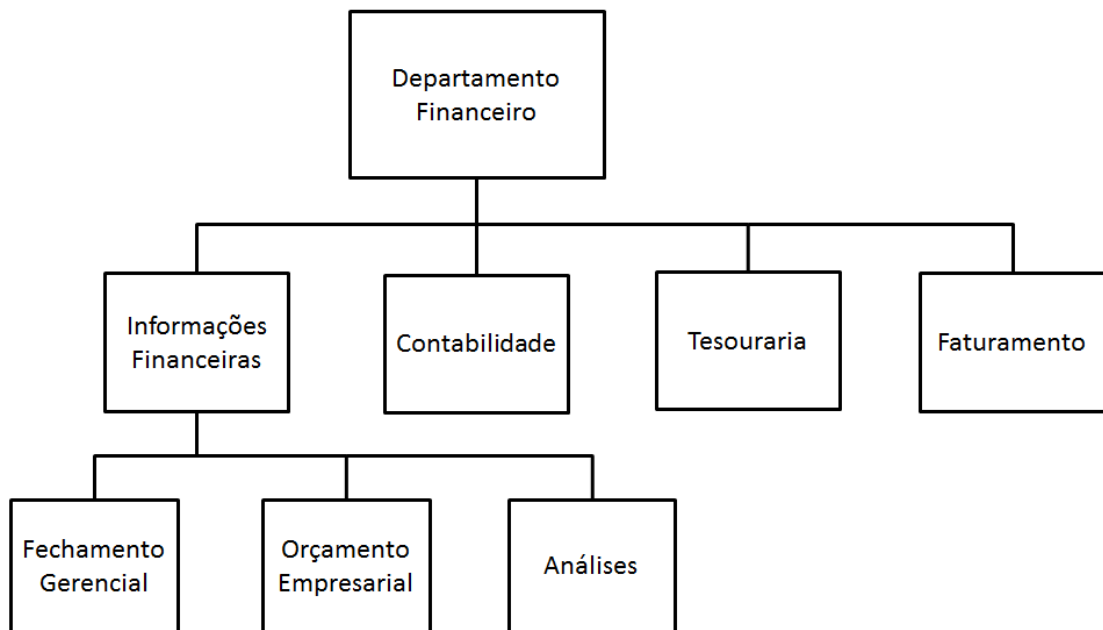


Figura 16 - Departamento financeiro da Empresa Beta
Fonte: Autoria do próprio autor.

Com a adoção dos conceitos de *data warehousing*, a expectativa da área era a de poder ter um grande histórico de informações disponíveis sobre o desempenho financeiro da organização, bem como poder comparar os dados gerados pelas suas duas equipes, isto é, conseguir comparar os valores orçados com os valores de fato realizados. Com essas informações em mãos, tanto a equipe da área quanto outras pessoas de outras áreas e regionais da empresa conseguem acompanhar o desempenho, fornecendo subsídios para o processo de tomada de decisão.

3.3 Visão geral do processo

Como visto na revisão bibliográfica, um *data mart* deve atender a apenas um processo de negócio da organização. No caso estudado, o processo escolhido para ser modelado e carregado no *data mart* está representado na Figura 17. A figura foi elaborada a partir das experiências vivenciadas pelo autor e pelo conhecimento adquirido sobre o processo durante o período em que passou na Empresa Beta.

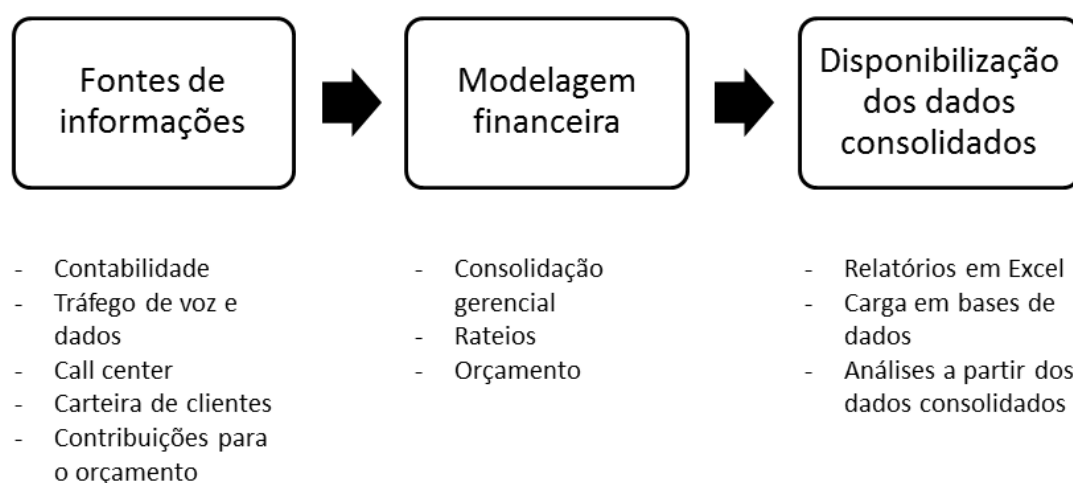


Figura 17 - Processo de negócio estudado

Fonte: Autoria do próprio autor.

O processo estudado começa com os dados provenientes de diversas fontes e que terão usos distintos durante as demais fases. Todos esses dados são provenientes de sistemas transacionais da empresa, como por exemplo os dados da contabilidade, que são extraídos do sistema ERP da empresa, com o saldo das diversas contas contábeis. Outros departamentos da organização fornecem dados sobre as ligações, a quantidade de minutos, a quantidade de interconexão com as outras operadoras, entre outras informações necessárias para a área. Outra importante fonte de informações são as contribuições para o orçamento, provenientes dos diversos departamentos e das diversas regionais da Empresa Beta. As contribuições são feitas via *browser* e ficam disponíveis para a ferramenta financeira.

Com os dados provenientes dessas diversas fontes, ocorre a modelagem financeira de todos os dados, de modo a fornecer dados consolidados que serão utilizados por gerentes e diretores

para basear suas decisões. Nessa etapa do processo, os dados provenientes principalmente da contabilidade são consolidados, pois os mesmos são recebidos no menor detalhe, isto é, por conta contábil e por centro de custo. Além da consolidação, é realizado o rateio dos custos e das despesas que possibilitem uma melhor visão gerencial sobre a performance financeira.

Outra atividade realizada na etapa da Modelagem Financeira é a de orçamento empresarial, que recebe contribuições das diversas áreas e regionais da empresa com previsões de gastos e de receitas para o ano corrente e para o ano seguinte, permitindo a consolidação desses dados e gerando um orçamento final, que será tomado como base para o acompanhamento financeiro ao longo dos meses.

A última etapa do processo analisado envolve a disponibilização dos dados gerados na etapa de Modelagem Financeira para serem acessados e utilizados por outras aplicações. Os dados podem ser carregados em bases de dados, como por exemplo no *data mart* do departamento financeiro, ou então também podem ser utilizados em planilhas Excel, tanto para geração de relatórios quanto para realizar variadas análises, como por exemplo analisar o impacto de possíveis investimentos, analisar o retorno de possíveis ações de marketing, entre outras possibilidades.

Dentre os dados disponibilizados pela área, estão Demonstrações do Resultado do Exercício (DRE), custos e despesas consolidados por regional, por produto, por diretoria, por mercado, relatórios do orçamento e comparação entre os dados orçados e os dados efetivamente realizados.

3.4 O projeto

A Empresa Beta contratou a Empresa Alfa no começo de 2010 para realizar um projeto de implantação de um sistema que servisse de suporte ao processo de tomada de decisão baseado em indicadores e dados financeiros. O sistema incluía uma ferramenta financeira e toda a modelagem dentro dessa aplicação, bem como um *data warehouse* que armazenasse os dados, criando uma base histórica.

Diante de mudanças internas da organização e alteração de requisitos por parte da matriz estrangeira da Empresa Beta, o sistema começou a perder aderência às necessidades da área de negócios, que começou a buscar alternativas para atender às suas necessidades urgentes, diminuindo assim a utilização do sistema de *data warehouse* então implantado.

Desse modo, a Empresa Beta voltou a contratar a Empresa Alfa para realizar um projeto de melhoria do sistema implantado, visando a adequação aos novos requisitos da área financeira e o resgate da utilização do *data warehouse*.

A equipe do projeto por parte da Empresa Alfa foi composta por um gerente de projeto, dois consultores sênior, um consultor júnior e um estagiário. O gerente tinha como funções a direção, coordenação e acompanhamento das atividades do projeto. Os consultores sênior e júnior tinham como atribuições levantar os requisitos junto aos clientes, analisar o cenário e propor uma solução, além de implementar a solução. O estagiário, no caso o autor do presente trabalho, era responsável por acompanhar todas as atividades dos consultores, auxiliando-os em qualquer uma de suas atribuições, além de ser responsável pela documentação da solução.

Os pontos de contato designados para o desenvolvimento do projeto por parte da Empresa Beta foram uma gerente da área de TI, uma analista da área de TI, dois gerentes da área financeira e três analistas da área financeira.

O projeto iniciou-se no mês de fevereiro de 2012 e terminou no mês de junho do mesmo ano, compreendendo 8 semanas de entendimento da situação, levantamento de requisitos, desenho e construção da solução, 3 semanas para realização de testes e homologação e 9 semanas de suporte pós-implantação. Após avaliação por parte da Empresa Beta, mostrou-se necessária a extensão do projeto, de modo a incluir melhorias específicas para a área de orçamento que tinham ficado de fora do escopo do primeiro projeto. Essa extensão teve início no mês de julho de 2012, com término no final de setembro do mesmo ano, envolvendo as etapas de levantamento de requisitos, desenho e construção da solução, testes, homologação e suporte. Nessa extensão do contrato, foram consideradas somente melhorias na etapa de Modelagem Financeira descrita no item 3.3 Visão Geral do Processo. O autor do trabalho ficou alocado a esse projeto durante toda a sua extensão.

3.5 Problemas

Os problemas levantados pela equipe da área financeira que levaram à contratação desse projeto junto à Empresa Alfa foram em primeiro lugar decorrentes da necessidade de readequação da solução previamente implantada, devido a mudanças internas da organização e à alteração de necessidades por parte da matriz estrangeira da Empresa Beta. Com isso, o sistema desenvolvido pela Empresa Alfa, bem como o *data warehouse*, começou a perder aderência com as reais necessidades da área de negócios. Para contornar esse problema, a área elaborou um processamento paralelo, com consolidação via Excel e carga em outro sistema de informações interno da empresa.

Esse processamento paralelo da área financeira demonstra a falta de aderência e de qualidade dos dados carregados no *data warehouse*. Devido à falta de confiabilidade dos dados carregados, o *data warehouse* passou cada vez mais a ser menos utilizado, com a crescente utilização da base de dados paralela. Outro problema levantado durante o período inicial do projeto foi o motivo da adoção dessa base de dados paralela. Internamente, o controle do processo estudado nesse caso é de responsabilidade da área de TI da Empresa Beta, cabendo à área financeira somente a execução do processo. Desse modo, alterações substanciais no processo passam por aprovações do departamento de TI, fato este que não agrada à área financeira. Assim, a área financeira desenvolveu o processamento paralelo e a base de dados paralela, que ficariam sob o seu controle.

3.6 Entendimento da situação anterior

O primeiro passo para poder propor alguma solução para o caso é entender a situação no seu quadro atual. Para isso, é essencial saber qual o ambiente no qual o caso se encontra, apresentando a organização, sua área específica de estudo, qual será o seu processo estudado e em que condições o trabalho foi desenvolvido. Todos esses pontos já foram descritos nos itens anteriores. Outro ponto de importância para o entendimento da situação é a definição e a análise dos aspectos mais técnicos e específicos desse caso. Assim, serão apresentados a seguir uma visão de como a macro estrutura foi desenvolvida durante o primeiro projeto em

2010, as análises segundo a revisão bibliográfica e por final as alterações realizadas para atender aos requisitos dos usuários.

3.7 Macroestrutura - Arquitetura

O primeiro aspecto a ser considerado durante o período inicial do projeto, de levantamento de requisitos e entendimento da situação inicial, foi levantar qual o fluxo de informações apresentado pelo processo de negócio de estudado. A partir do entendimento do autor deste trabalho sobre o processo da área financeira, foi elaborada a Figura 18 contendo todas as etapas da cadeia de informações. Foi utilizado o modelo de Kimball para representar os diversos componentes do ambiente do *data warehouse*.

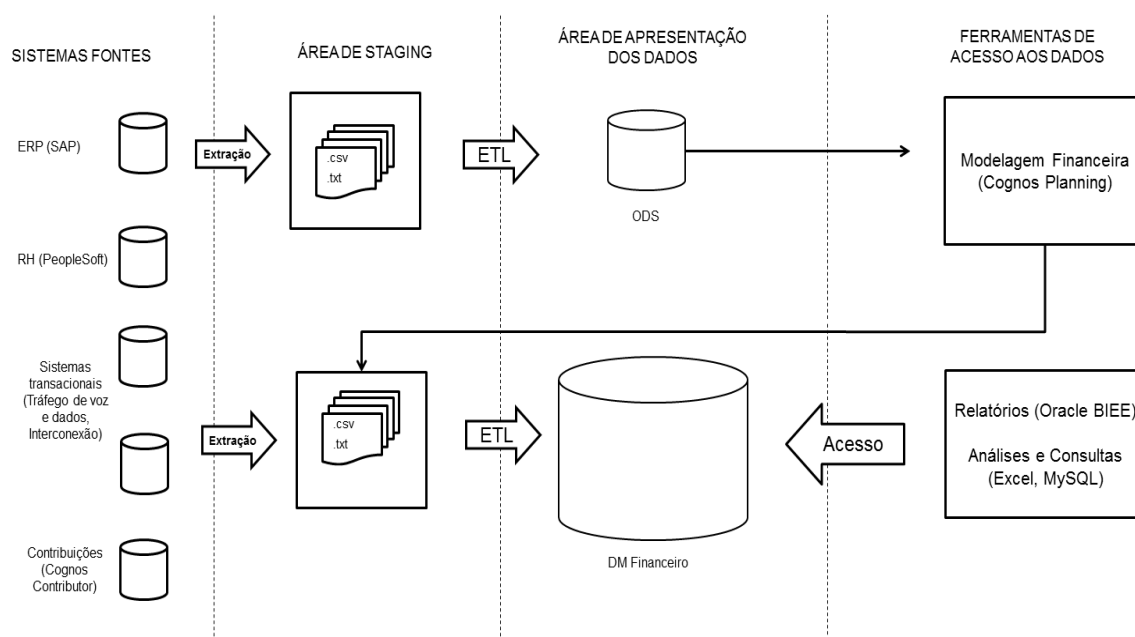


Figura 18 - Macro estrutura do ambiente do *data warehouse* da área financeira da Empresa Beta
Fonte: Autoria do próprio autor.

3.7.1 Sistemas Fontes

Os sistemas fontes são aqueles que fornecem dados para o processo de negócio estudado. Foram identificados as seguintes fontes de dados:

- Sistema ERP

O sistema ERP adotado pela Empresa Beta é o fornecido pela SAP, líder de mercado no fornecimento de softwares do gênero. Para o processo de Modelagem Financeira, os dados da contabilidade são extraídos do sistema ERP, gerando arquivos do tipo “.csv”² com os saldos de todas as contas contábeis por centros de custo.

- Sistema de RH

O sistema de Recursos Humanos utilizado pela Empresa Beta para controlar todos os aspectos relacionados à gestão de pessoas é o PeopleSoft, que administra o cadastro dos funcionários, fornecendo informações sobre a folha de pagamento, a pontualidade, entre outros recursos. Os dados provenientes do software PeopleSoft são utilizados somente para o orçamento, para o cálculo da folha consolidada por regional e por centro de custo.

- Sistemas transacionais (tráfego de voz e dados, interconexão)

Os dados referentes às funções operacionais de uma empresa de telefonia móvel, como por exemplo registro de chamadas, tráfego de voz e dados e interconexão com outras operadoras abastecem o processo financeiro fornecendo *drivers* que serão utilizados para os rateios gerenciais. Os dados são fornecidos pela área de Engenharia da Empresa Beta, que os extrai de seus sistemas transacionais e os disponibiliza em arquivos de texto para a área financeira.

- Sistema de contribuições

A principal fonte de informação para o processo de orçamento são as contribuições das diversas áreas e regionais com suas expectativas de gastos para o decorrer do ano corrente e para todo o ano seguinte. Essas previsões de gastos são inseridas na ferramenta de contribuições da empresa, conhecida como Cognos Contributor. O Cognos Contributor é um módulo da ferramenta de modelagem financeira adotada pela Empresa Beta, o Cognos

² Comma-separated values: Valores Separados por vírgulas. Formato de arquivo bastante utilizado para representar dados tabelados.

Planning. O Cognos Contributor possui interface amigável com o usuário, pois pode ser acessado diretamente via *browser*, facilitando a questão da acessibilidade.

3.7.2 Área de *Staging*

De acordo com o item 2.6.2.1 – Componentes de um *Data Warehouse* segundo Kimball, a Área de *Staging* não deve ser acessível para os usuários finais, sendo uma etapa anterior ao processo de carga nas bases. A Empresa Beta segue esta recomendação de Kimball, protegendo a Área de *Staging* contra acessos indevidos.

É na Área de *Staging* que ocorre o processo de ETL, onde os dados são transformados de diversas maneiras, principalmente com o intuito de torná-los coerentes entre si e coerentes com as tabelas do *data warehouse*, e finalmente são carregados na Área de Apresentação dos Dados, na ODS ou no *data mart* financeiro.

Todos os dados que serão utilizados pela ferramenta de modelagem financeira são carregados na ODS. Os dados já modelados pela ferramenta financeira Cognos são carregados no *data mart* financeiro juntamente com alguns dados dos sistemas transacionais.

A *Operational data store* (ODS) foi escolhida para ser utilizada como uma etapa anterior à carga dos dados no software de modelagem financeiro Cognos para facilitar o processo dentro da aplicação Cognos. Devido ao tamanho elevado dos arquivos e do número de linhas desses arquivos, a adoção da ODS permitiu simplificar o processo de carga dentro da ferramenta ao mesmo tempo que tornou o processo mais rápido. A velocidade de acesso a tabelas de dados é bem maior do que a velocidade para acessar as informações disponíveis em arquivos de texto ou do tipo “.csv”.

O software adotado pela Empresa Beta para o processo de ETL é o Informatica PowerCenter, responsável por administrar e processar todo o fluxo de informações que será carregado nas bases da ODS e do *data mart* financeiro. O desenvolvimento das lógicas do processo é feito através da utilização de mapas, que indicam de onde vêm os dados, quais são os processos de transformação pelos quais passam e onde devem ser carregados.

3.7.3 Área de Apresentação dos dados

A Área de Apresentação dos dados é de fato onde os dados são armazenados fisicamente e, portanto, os usuários finais da área de negócios consideram esse ambiente como sendo o *data warehouse*. No caso da Empresa Beta, a Área de Apresentação dos dados apresenta dois componentes: a *Operational Data Store* (ODS) e o *data mart* financeiro.

A ODS é utilizada para armazenar os dados provenientes dos diversos sistemas transacionais da empresa, como o SAP, o PeopleSoft e os sistemas de voz e dados, após passar por um processo de integração na etapa anterior. Os dados contidos na ODS irão alimentar o software de modelagem financeira adotado pela Empresa Beta. A ODS somente é utilizada para o processo da equipe de fechamento gerencial.

O outro componente do ambiente de *data warehouse* é o DM Financeiro, ou então, *data mart* financeiro. Ele é a base de dados departamental responsável por armazenar as informações resultantes do processo de modelagem financeira, que ficarão disponíveis para as ferramentas de acesso aos dados, como por exemplo geradores de relatórios, ferramentas analíticas ou então para simples consultas.

3.7.4 Ferramentas de acesso aos dados

A camada final do ambiente estudado é o das Ferramentas de acesso aos dados. Nessa camada, se encontram todas as aplicações que acessam os dados contidos na Área de Apresentação dos dados. No caso da Empresa Beta, as ferramentas podem acessar tanto a ODS quanto o *data mart* financeiro. Como apresentado na Figura 18, somente a ferramenta de modelagem financeira Cognos Planning acessa a ODS, enquanto que diversas ferramentas podem acessar o *data mart* financeiro, como o gerador de relatórios Oracle BIEE, a ferramentas de consultas PL/SQL, o Microsoft Excel e inclusive o próprio Cognos Planning.

IBM Cognos® 8 Planning - Analyst

File Edit D-Cube Window Help

[D Cube] Fechamento Gerencial.11 Opex

COT10000 Financeira Reg 201103 RLZ V-01 Original

	Original	Ajuste	Total	Opex_Investimento	Detalhe Opex de Gestão/Investimento no Cliente	DIRETORIA_NIVEL_1	Detalhe Investimento	CR VENDAS CONSUMO
3141010204 DOACOES NAO DEDUTIVEIS NA - 04	0	0	0	Opex de Gestio		Financas		
3121109999 DESPESAS COM DOACOES E CORTESIAS	0	0	0	Opex de Gestio		Financas		
3141010203 AJUSTE CARGA DE DOACOES D - 03	0	0	0	Opex de Gestio		Financas		
Folha	63.006	0	63.006					
Alimentação - Pat	3.582	0	3.582					
3121010504 ALIMENTACAO PAT	3.582	0	3.582	Opex de Gestio		Financas		
Assistência Médica E Odontológica	2.930	0	2.930					
3121010502 ASSISTENCIA MEDICA	2.818	0	2.818	Opex de Gestio		Financas		
3121010503 ASSISTENCIA ODONTOLOGICA	112	0	112	Opex de Gestio		Financas		
Estatários	0	0	0					
3121010506 BOLSA AUXILIO	0	0	0	Opex de Gestio		Financas		
Férias	5.581	0	5.581					
3121010103 ABONO PECUINIARIO	0	0	0	Opex de Gestio		Financas		
3121010307 FERIAS	5.581	0	5.581	Opex de Gestio		Financas		
PGRS	3.311	0	3.311					
3121010306 FUNDO DE GARANTIA POR TEM - 06	3.311	0	3.311	Opex de Gestio		Financas		
Qualificação	0	0	0					
3121010205 AJUDA DE CUSTO	0	0	0	Opex de Gestio		Financas		
3121010105 ADICIONAL DE TRANSFERENCIA	0	0	0	Opex de Gestio		Financas		
3121010201 OUTRAS GRATIFICACOES	0	0	0	Opex de Gestio		Financas		
3121010204 BONUS	0	0	0	Opex de Gestio		Financas		
Hora Extra	2.009	0	2.009					
3121010102 HORA EXTRA	2.009	0	2.009	Opex de Gestio		Financas		
3121010107 HORAS EXTRAS SOBRE AVISO - 07	0	0	0	Opex de Gestio		Financas		
Indenização Trabalhista	9	0	9					
3121010406 PROVENTOS DIVERSOS	9	0	9	Opex de Gestio		Financas		
3121010401 AVISO PREVIO	0	0	0	Opex de Gestio		Financas		
3121010402 INDENIZACAO TRABALHISTA	0	0	0	Opex de Gestio		Financas		
3121010410 PDV INDENIZACAO TRABALHISTA	0	0	0	Opex de Gestio		Financas		
3121010411 MULTA S RESCISOES	0	0	0	Opex de Gestio		Financas		
INSS	8.278	0	8.278					
3121010301 PREVIDENCIA SOCIAL	8.278	0	8.278	Opex de Gestio		Financas		
Outros Benefícios	325	0	325					
3121010407 AJUDA AO EXCEPCIONAL	0	0	0	Opex de Gestio		Financas		
3121010403 COMPLEMENTO DE AUXILIO DOENCA	0	0	0	Opex de Gestio		Financas		
3121010404 AUXILIO EDUCACAO	0	0	0	Opex de Gestio		Financas		
3121010408 CRECHE AMAMENTACAO	325	0	325	Opex de Gestio		Financas		
3121010522 DESCONTO FARMACIA	0	0	0	Opex de Gestio		Financas		
3121010523 CONVENIO FACULDADE	0	0	0	Opex de Gestio		Financas		
3151010523 PROVIDORES CONVENIO FACULDADE	0	0	0	Opex de Gestio		Financas		
Outros Encargos	3.485	0	3.485					
TOTAL DESPESA								

Iniciar

Mensagens Recebidas...

p_cognos em "paulo..."

IPMB Cognos® 8 Pla...

Documento1 - Micros...

PT

11:13

Figura 19 - Exemplo de tela do Cognos Planning da Empresa Beta

IFRS									
Estado de Resultados Institucional Consolidado									
(Milhares de Reais)									
Período: JANEIRO de 2012									
RLZ	M-1 RLZ	BDO	M-12 RLZ	Conceitos	RLZ YTD	RLZ M-12 YTD	BDO YTD		
\$	\$	\$	\$	YOY	\$	\$	\$	YOY	
292.054		0		TOTAL DESPESAS	292.054				100.0
137.879		0		Custo de Serviços	137.879		0		100.0
214		0		Interconexão	214		0		100.0
1.033		0		Cobilling	1.033		0		100.0
9.918		0		Rearranging	9.918		0		100.0
90		0		Detrat	90		0		100.0
908		0		Portabilidade	908		0		100.0
14.045		0		Provedores	14.045		0		100.0
68.703		0		Fiscal Manutenção	68.703		0		100.0
20.735		0		Frais Habilitação	20.735		0		100.0
4.387		0		I Circuitos de dados	4.387		0		100.0
15.254		0		Site Expenses	15.254		0		100.0
11.410		0		OSM	11.410		0		100.0
1.889		0		Logística e Armazenagem de Engenharia	1.889		0		100.0
287		0		Impressão de Cartões	287		0		100.0
17.686		0		Custo de Operações	17.686		0		100.0
9.287		0		Billing	9.287		0		100.0
3.718		0		Armazenagem	3.718		0		100.0
3.429		0		Cobrança	3.429		0		100.0
2.271		0		Crédito	2.271		0		100.0
50.893		0		Subsídio or Impacto Pro-DIF	50.893		0		100.0
4.338		0		Pro-DIF	4.338		0		100.0
55.232		0		Subsídio	55.232		0		100.0
29.541		0		Marketing Cliente	29.541		0		100.0
33.249		0		Advertising	33.249		0		100.0
-5.034		0		Rebate	-5.034		0		100.0
1.335		0		Fidelidade	1.335		0		100.0
-10		0		Pesquisa mercadológica	-10		0		100.0
74.992		0		Comissões	74.992		0		100.0
56.381		0		Comissão	56.381		0		100.0
14.215		0		Propaganda Cooperada	14.215		0		100.0
4.366		0		Investimento em Agente Autorizado	4.366		0		100.0
90.338		0		Outros Investimentos no Cliente	90.338		0		100.0
31.073		0		Aprendimento	31.073		0		100.0
2.693		0		Tele vendas	2.693		0		100.0
12.030		0		Promotores de Vendas	12.030		0		100.0
21.921		0		Lógica	21.921		0		100.0
3.457		0		RCV	3.457		0		100.0
8.362		0		Vendas Corporativas	8.362		0		100.0
10.802		0		Logística e Armazenagem de Aparelhos	10.802		0		100.0
24.541		0		PDO Total	24.541		0		100.0
24.990		0		PDO Serviços	24.990		0		100.0
-449		0		PDO Aparelhos	-449		0		100.0
166.185		0		OPEX DE GESTÃO	166.185		0		100.0

Figura 20 – Exemplo de tela do Oracle BIEE

3.8 Análises

Apresentada a macro estrutura do ambiente do *data warehouse* da Empresa Beta, é possível realizar algumas análises tendo como base a revisão bibliográfica bem como apresentar as análises relativas ao âmbito do projeto.

3.8.1 Análises segundo a revisão bibliográfica

As análises apresentadas a seguir nesta seção foram elaboradas pelo autor do trabalho, utilizando do conhecimento adquirido durante o período que passou alocado na Empresa Beta, bem como tendo como base os itens levantados na revisão bibliográfica.

3.8.1.1 Análise da arquitetura

Analisando a arquitetura apresentada na Figura 18, pode-se afirmar que foi utilizada a abordagem de Kimball para a construção do ambiente de *data warehouse*. Os sistemas transacionais são os sistemas fonte para o processo, que após passarem pelo software de modelagem financeira Cognos Planning e por um processo de ETL, são carregados diretamente no *data mart* departamental da área financeira. Não existe um grande repositório de dados único no caso da Empresa Beta, do qual poderiam ser criadas as bases departamentais conhecidas como *data marts*. A situação está exemplificada na Figura 21.

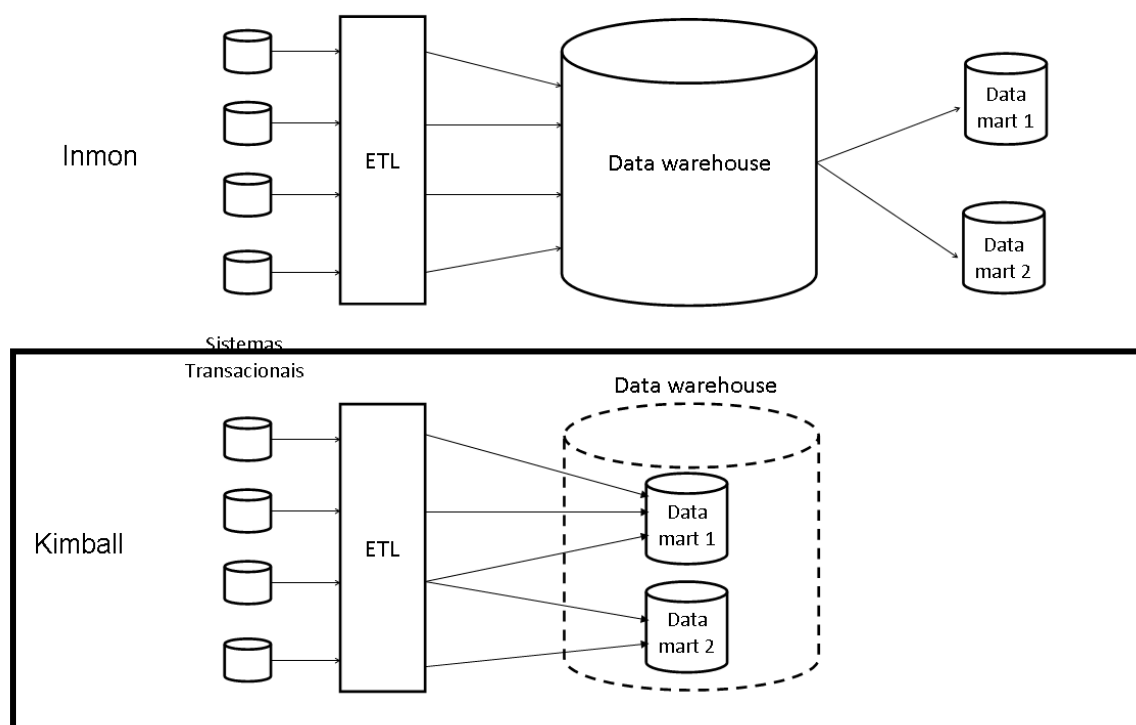


Figura 21 - Análise da arquitetura implantada

Fonte: Autoria do próprio autor.

De acordo com o item 2.8 Escolhendo a melhor abordagem, são oito os critérios a serem utilizados para a avaliação de qual abordagem de *data warehouse* se enquadra melhor em cada caso, a de Inmon ou a de Kimball. Como visto anteriormente, a abordagem adotada para a Empresa Beta parece estar mais alinhada aos pensamentos de Kimball. Serão utilizados os oito critérios para confirmar que foi feita a melhor escolha para as necessidades da área financeira e da Empresa Beta.

Característica	Favorece Kimball	Favorece Inmon
Natureza das decisões	Táticas	Estratégicas
Requisitos de integração de dados	Áreas de negócios individuais	Integração em toda a empresa
Estrutura dos dados	Métricas de negócio, medidas de desempenho, scorecards	Dados que não são métricas; dados que devem atender diferentes e variadas necessidades
Escalabilidade	Necessidade de adaptação a requisitos altamente voláteis, dentro de um escopo limitado	Escopo crescente e requisitos mutantes são fatores críticos
Persistência dos dados	Sistemas fonte são relativamente estáveis	Sistemas fonte possuem alta taxa de mudanças
Requisitos de habilidades e pessoal	Pequenas equipes de generalistas	Grande equipe(s) de especialistas
Tempo para entrega	Necessidade da primeira aplicação de data warehouse é urgente	Requisitos da organização permitem um tempo maior até a primeira aplicação
Custo de implantação	Baixo custo inicial, com projetos subsequentes custando aproximadamente o mesmo	Alto custo inicial, porém com menores custos para projetos subsequentes



Enquadramento da Empresa Beta

Figura 22 - Análise da escolha da melhor abordagem de *data warehouse*

Fonte: Autoria do próprio autor.

Como pode ser visto na Figura 22, o enquadramento do caso da Empresa Beta tende a ser mais voltado para a abordagem de Kimball. As decisões que as informações contidas no *data warehouse* pretendem suportar são tanto de ordem tática quanto de ordem estratégica, pois podem ser utilizados tanto os dados consolidados, como os apresentados em DREs, que em geral tem um público mais estratégico, quanto as informações consolidadas de uma certa regional ou de um certo centro de custo, com um público mais tático.

Os dados a serem armazenados são métricas de negócios e medidas de desempenho, como as informações fornecidas por um DRE, como por exemplo o total de receita, custos de

operação, custos de comercialização, Lucros antes de juros, impostos, depreciação e amortização (LAJIDA, ou EBITDA em inglês), % LAJIDA, entre outros.

Há a necessidade de adaptação a requisitos altamente voláteis, como por exemplo uma nova determinação da ANATEL, de outro órgão público ou até mesmo da matriz estrangeira, que obrigaria mudanças na forma de reportar os dados ou de medir as métricas de negócio em um curto prazo. Apesar dessa necessidade de adaptação rápida, o escopo dessas mudanças é limitado, pois compreende somente os assuntos referente ao processo específico atendido por esse *data mart*.

Os sistemas fonte são relativamente estáveis, não havendo mudanças frequentes e de grande impacto na formatação dos dados. Também não é muito frequente a mudança dos sistemas fonte do processo, pois se tratam de sistemas consolidados dentro da organização, como o ERP, o sistema de RH e os sistemas transacionais de registro de ligações e operações.

Na época de implantação do ambiente de *data warehouse*, no projeto desenvolvido no ano de 2010, existia certa urgência pela primeira aplicação de *data warehouse*, para atender ao orçamento que já se encontrava em processo. Também foi considerado o custo inicial de implantação neste caso. A Empresa Beta possuía um orçamento apertado, suficiente apenas para a implantação do *data mart* da área financeira, deixando os outros *data marts* para serem implantados em projetos futuros.

3.8.1.2 Análise da ODS

De acordo com a Figura 5 do item 2.4 *Operational Data Store* (ODS), a ODS deve ser utilizada como uma etapa anterior à carga dos dados no ambiente de *data warehouse*. Observa-se que foi utilizada uma abordagem diferente para a ODS no caso da Empresa Beta. A ODS foi implementada para servir ao software de modelagem financeira. Assim, foi introduzida uma etapa intermediária entre a base ODS e a base do *data warehouse*. A Figura 23 demonstra o caso descrito.

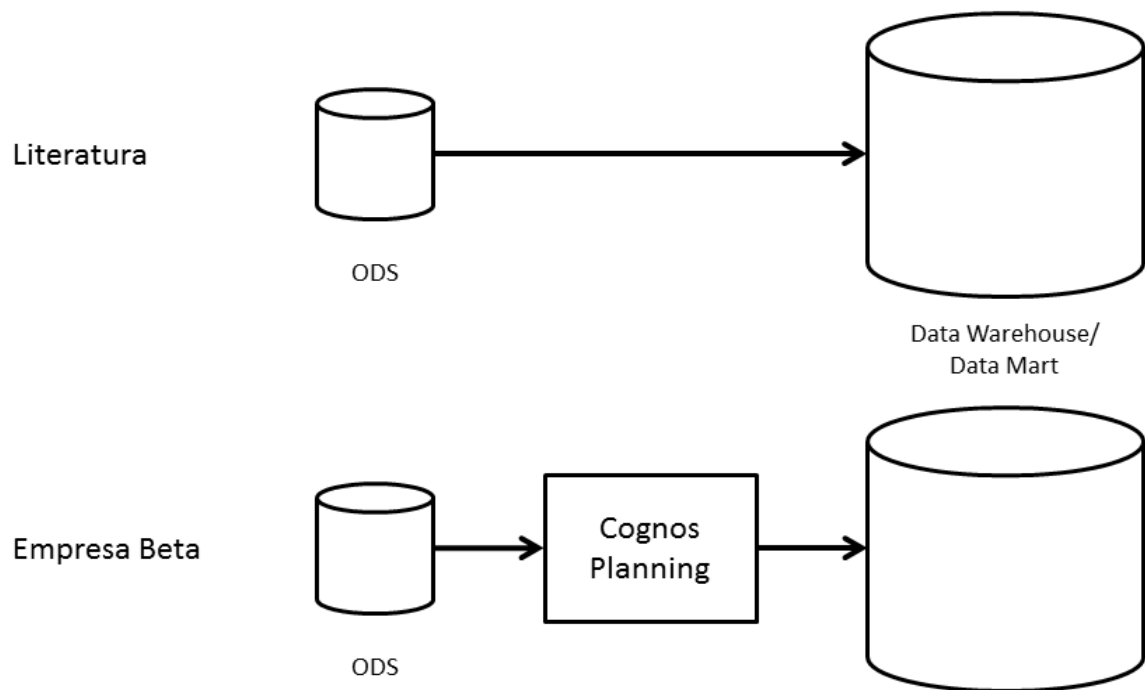


Figura 23 - Análise ODS
Fonte: Autoria do próprio autor.

Essa extensão do sistema de *data warehouse* para os sistemas transacionais pode ser explicada pela necessidade de integração dos dados dos diversos sistemas fonte “legados” antes de serem carregados para o software de modelagem financeira. Ao adotar o conceito de ODS, a empresa consegue um alívio imediato, diminuindo o tempo de processamento e carga dos dados, pois os dados seguirão o mesmo padrão, não necessitando de processos específicos para cada fonte de dados. No caso estudado, o processamento diminuiu de 6 horas de duração para apenas 20 minutos depois de adotado o conceito de ODS para a carga no Cognos Planning.

3.8.2 Análises do projeto

Foram realizadas entrevistas com os pontos focais da área financeira da Empresa Beta para o levantamento dos requisitos e das necessidades. Participou por parte da Empresa Alfa toda a equipe, desde o gerente do projeto até o estagiário.

A partir das conversas com os usuários da área de negócios, ficou evidente a necessidade de readequação da solução previamente implantada. O grupo controlador estrangeiro mudou o formato pelo qual os dados devem ser reportados, impactando a solução desenvolvida em 2010. Com essa mudança no formato de reportar os dados da área financeira, a equipe teve que desenvolver uma consolidação paralela em Excel e utilizar uma outra base de dados corporativa existente para carregar suas informações, pois o processo de modelagem financeira ficou defasado e a equipe não possuía recursos para adaptar o processo às novas necessidades.

Com base nesse cenário, a equipe de projeto começou a analisar quais seriam os impactos dessa mudança nos componentes da solução previamente implantada. A Tabela 8 apresenta os componentes do modelo e o seu diagnóstico.

Componente	Situação
Sistemas Fontes	Sem alterações. Mesmas fontes de informações.
Área de Staging	Processo de ETL para a ODS inalterado. Processo de ETL para o DM Financeiro precisará de alterações.
Área de Apresentação dos dados	ODS inalterada. Alterações necessárias no DM Financeiro.
Ferramentas de acesso aos dados	Mudanças na modelagem financeira no software Cognos Planning. Alterações no gerador de relatórios Oracle BIEE.

Tabela 8 - Análise das alterações necessárias na solução implantada

As mudanças na modelagem financeira, de modo a adaptar-se às novas necessidades, provocam, como consequência, as alterações no modelo do *Data Mart* Financeiro, nos modelos dimensionais (*Star Schema*), nas tabelas físicas e no processo de ETL para carregar o *data mart*.

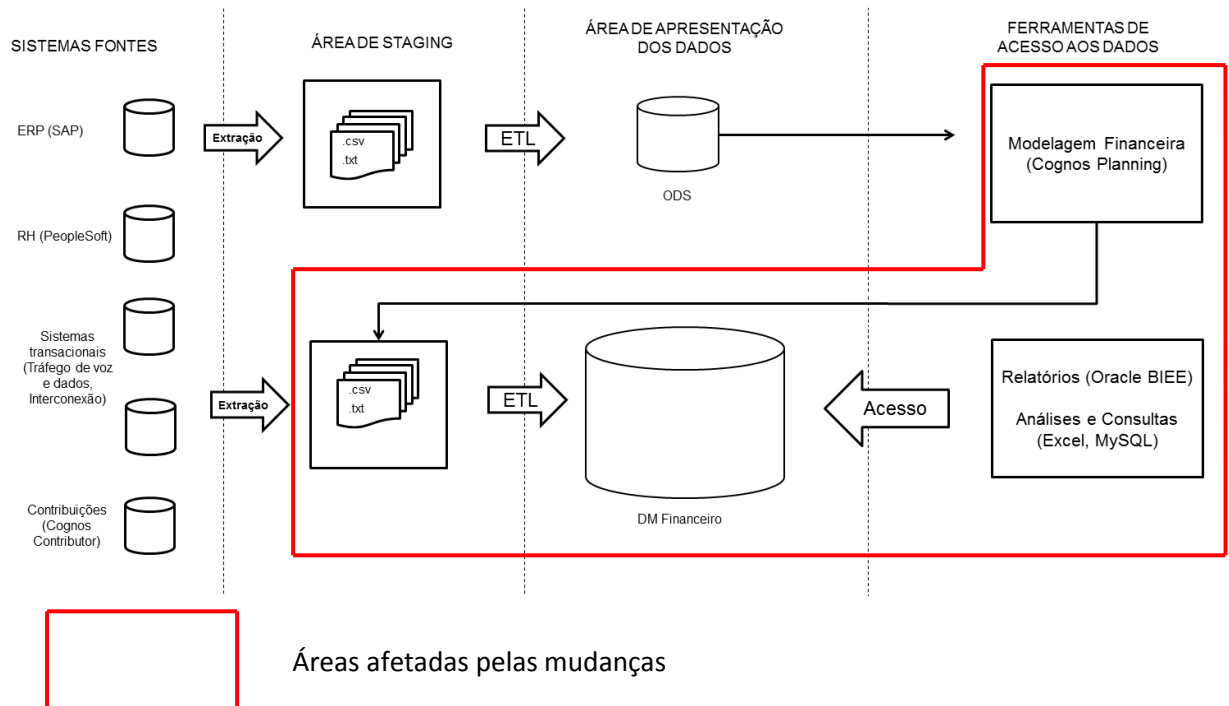


Figura 24 - Análise dos componentes afetados pela readequação

3.8.2.1 Modelagem dimensional

Como levantado anteriormente, ocorreram mudanças na camada de modelagem financeira, de modo a adaptar o Cognos Planning às novas necessidades de formatação dos dados e dos relatórios. Essas mudanças tem impacto direto no *data mart* financeiro e em como ele foi modelado.

De maneira geral, foi necessária a inclusão de uma nova dimensão nessas tabelas, de modo a comportar os dados provenientes tanto na nova formatação quanto na formatação antiga. As duas visões dos dados são necessárias, pois elas apresentam diferenças de granularidade e de apresentação.

As tabelas fato identificadas que deveriam sofrer alterações foram as dos seguintes assuntos: Custo Despesa, Custo Opex e DRE. A tabela de Custo Despesa apresenta os valores de despesas já consolidados em alguns agrupadores, ao passo que a de Custo Opex apresenta os valores das contas contábeis por centros de custo. A tabela de DRE é relativa aos dados

gerados no formato de Demonstrações do Resultado do Exercício, como receita, custos de operação, custos de comercialização, despesas financeiras, etc. As novas dimensões criadas foram as relativas ao país, à diretoria do grupo controlador e à visão do DRE.

A seguir são apresentados os modelos dimensionais das tabelas fato que sofreram alterações, com as modificações destacadas em vermelho. A indicação (FK) diz respeito a chaves estrangeiras à tabela fato.

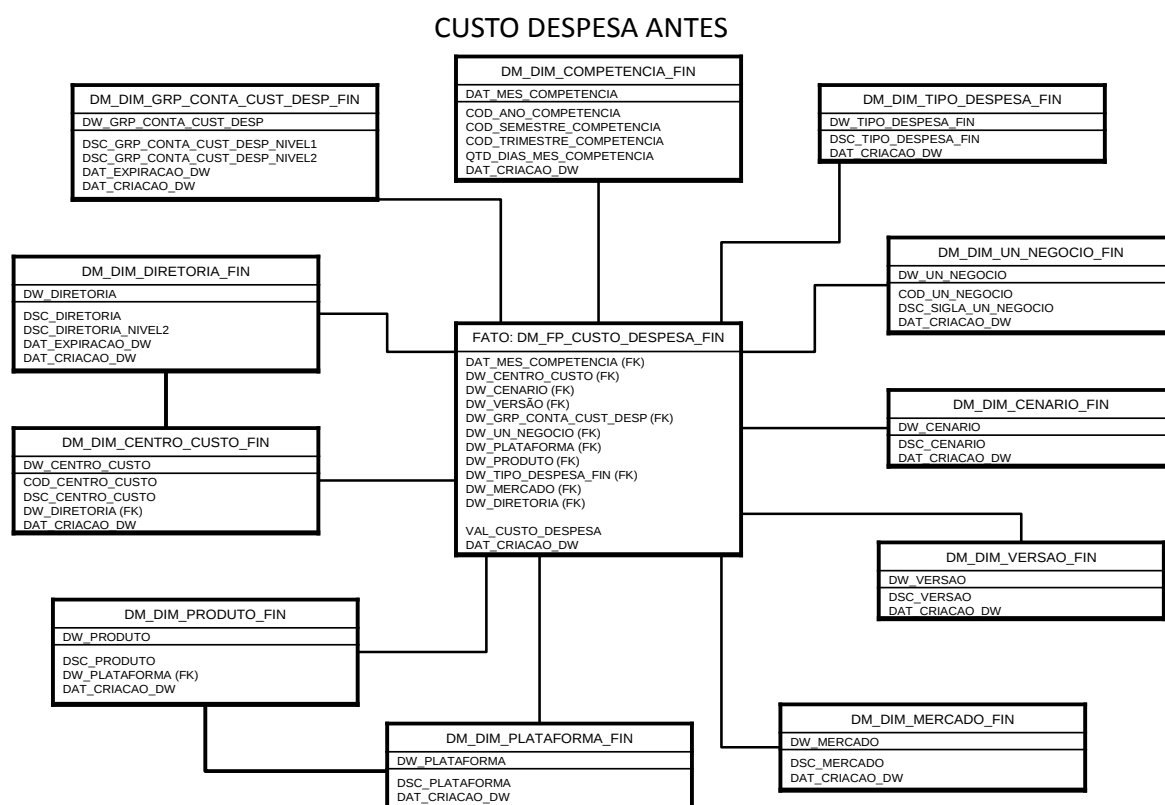


Figura 25 - Fato de Custo Despesa antes do projeto

CUSTO DESPESA DEPOIS

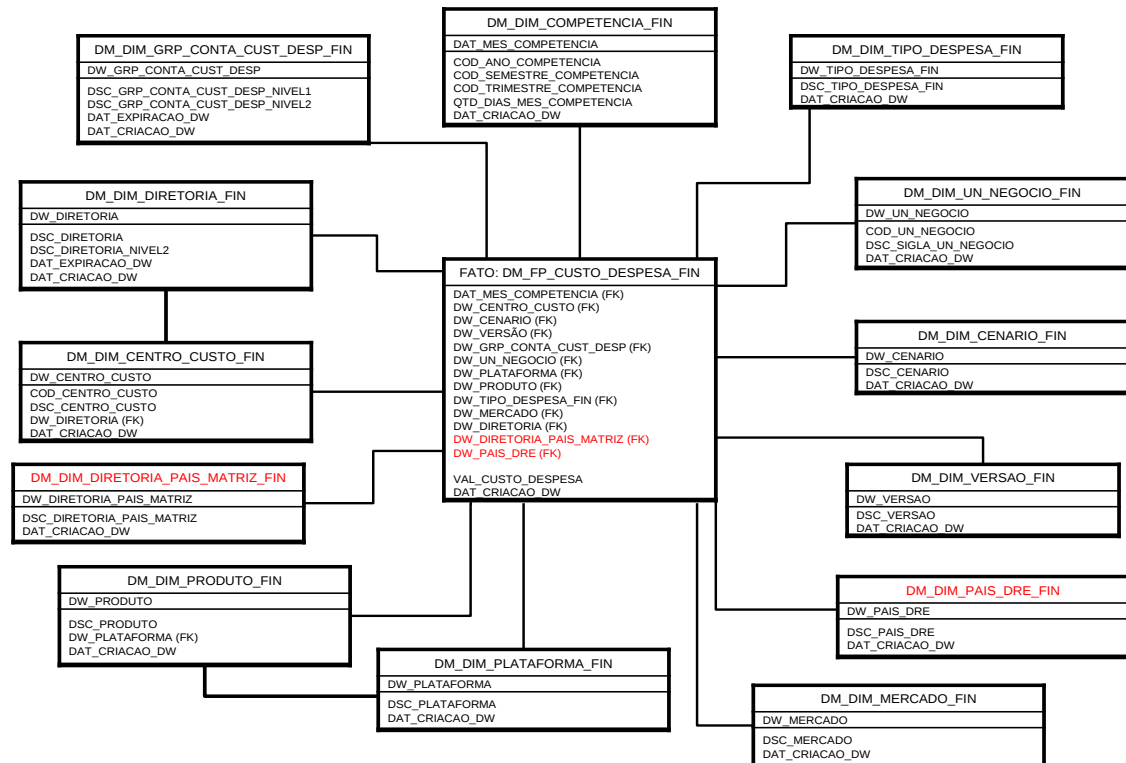


Figura 26 - Fato de Custo Despesa depois do projeto

CUSTO OPEX ANTES

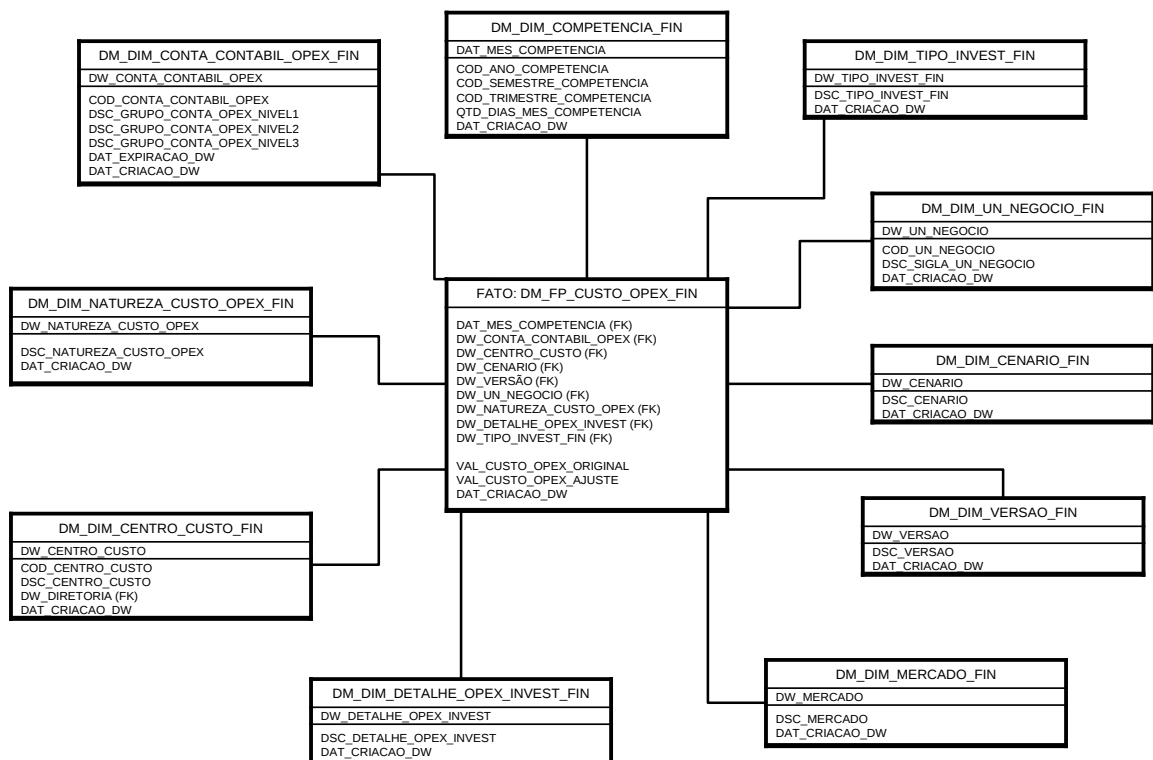


Figura 27 - Fato de Custo Opeex antes do projeto

CUSTO OPEX DEPOIS

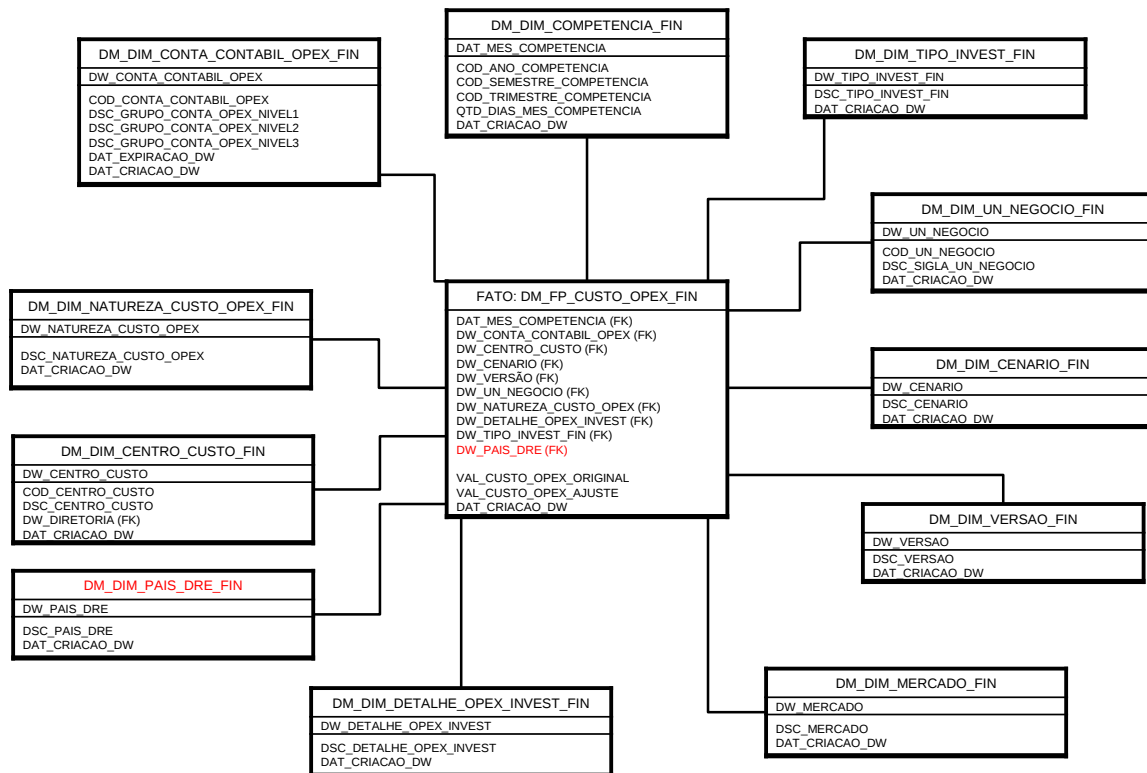


Figura 28 - Fato de Custo OpeX depois do projeto

DRE ANTES

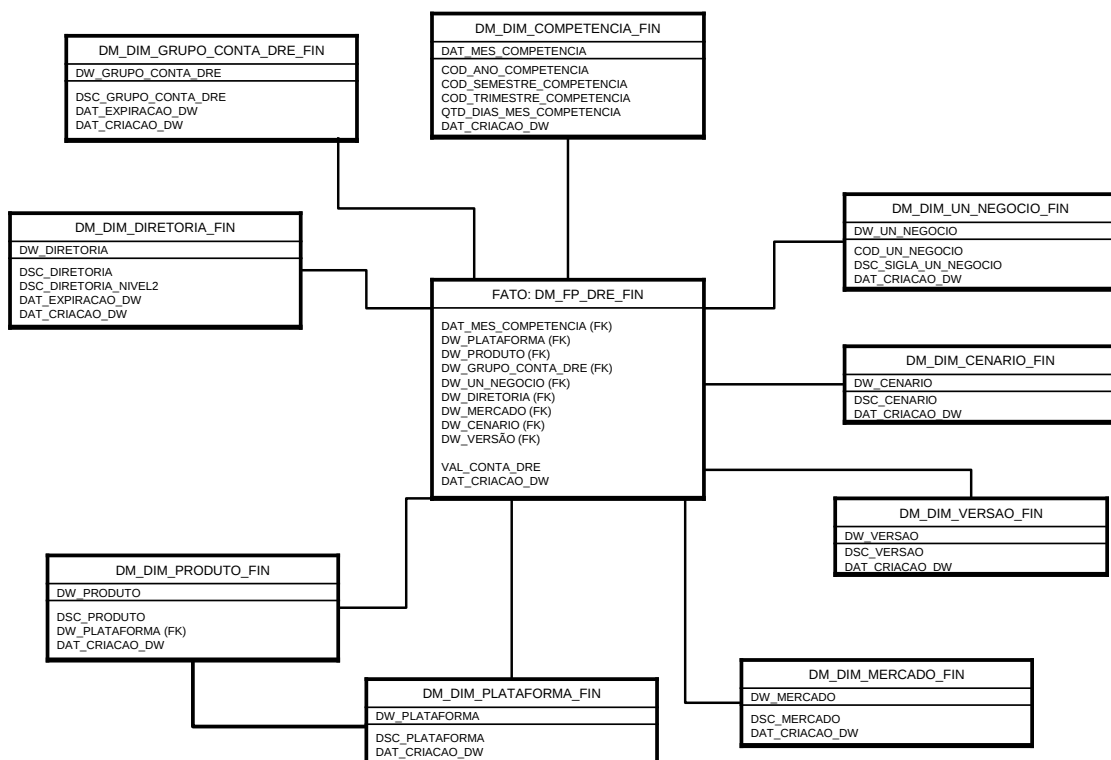


Figura 29 - Fato de DRE antes do projeto

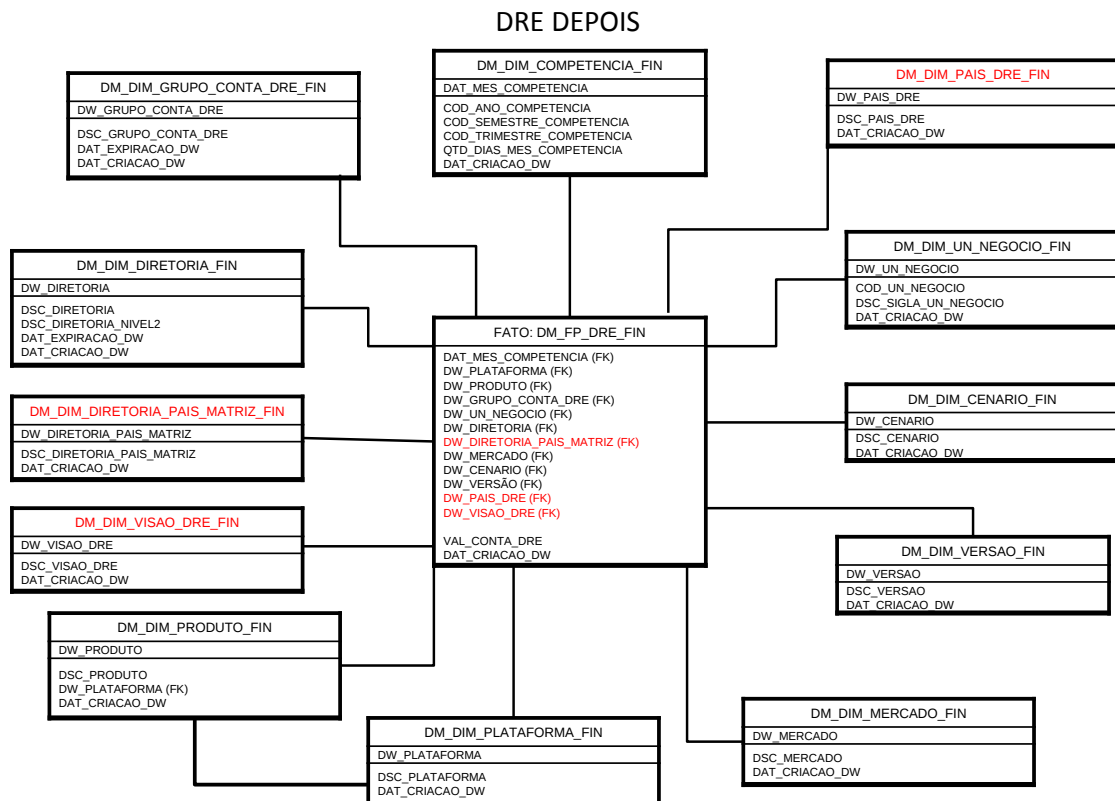


Figura 30 - Fato de DRE depois do projeto

Observando as tabelas fato acima, nota-se a inexistência de chaves primárias (PK) nas fatos. As chaves primárias estão presentes somente nas dimensões, se tornando chaves estrangeiras dentro da tabela da fato. A inclusão de uma chave primária nas tabelas permitiria encontrar um registro somente pelo número de sua chave, que é único. Entretanto, essa chave extra aumentaria o número de colunas e portanto o tamanho físico ocupado pelas tabelas. Essa ausência de chaves primárias nas fatos não prejudica o modelo dimensional, pois a combinação das diversas chaves estrangeiras formam entradas únicas de dados.

Outro fato com relação às tabelas fato apresentadas acima está na existência de dimensões que tem ligações com outras dimensões, como é o caso das dimensões de diretoria e de centro de custo. Esse fato demonstra a desnormalização apresentada pela utilização da modelagem dimensional.

Com relação ao aspecto físico do modelo e das tabelas, foram criadas as seguintes tabelas dentro do *data mart*: DM_DIM_DIRETORIA_PAIS_MATRIZ_FIN, DM_DIM_PAIS_DRE_FIN e DM_DIM_VISAO_DRE_FIN. As seguintes tabelas foram

alteradas de modo a incluir os campos para as novas dimensões: DM_FP_CUSTO_DESPESA_FIN, DM_FP_CUSTO_OPEX_FIN e DM_FP_DRE_FIN.

O processo de ETL foi modificado somente para a aplicação do *data mart* financeiro, de modo a considerar a carga das novas tabelas de dimensões, bem como considerar a inclusão dessas dimensões nas informações das fatos.

3.8.2.2 Qualidade dos dados

A qualidade dos dados contidos no *data warehouse* foi um dos problemas identificados durante a etapa de levantamento de requisitos juntos aos clientes. Devido à necessidade de mudanças no formato como os dados financeiros são reportados para a matriz estrangeira, o processo claramente se tornou fraco e obsoleto. Entretanto, antes mesmo dessa necessidade de mudança, o autor desse projeto presenciou, através de conversas informais com analistas da área financeira, a desconfiança e falta de credibilidade dos dados do *data warehouse*.

De acordo com Orr (1998), a qualidade dos dados em um sistema de informações é função de seu uso, e não de sua coleção. No caso da Empresa Beta, foi levantado que o *data warehouse* não estava sendo muito utilizado, com a área financeira se utilizando inclusive de bases paralelas ao *data warehouse*, fato este que demonstra a baixa qualidade de dados.

De acordo com a revisão apresentada no item 2.9 Qualidade dos dados do *Data warehouse*, são quatro as dimensões a serem consideradas ao analisar a qualidade de dados em um sistema informacional. A Tabela 9 apresenta o diagnóstico para o caso da Empresa Beta.

Dimensão	Situação da Empresa Beta
Acuracidade	Problemas de acuracidade levantados pela área de negócio. Divergência entre dados de origem e os dados contidos no <i>data warehouse</i> .
Disponibilidade	A informação está disponível para ser consultada.
Atualização	A informação está atualizada, mas apresenta problemas

	de confiabilidade.
Confiabilidade/Credibilidade	Os dados contidos no <i>data warehouse</i> não representam as necessidades da área de negócios.

Tabela 9 - Dimensões da qualidade de dados aplicadas na Empresa Beta
Fonte: A autoria do próprio autor.

De modo a combater a baixa qualidade de dados em cada uma das dimensões, foram elencadas algumas medidas, focadas principalmente em incentivar o uso dos dados por parte dos usuários da área de negócios. As medidas propostas para cada dimensão estão representadas na Tabela 10.

Dimensão	Medidas
Acuracidade	Grande período de suporte Realização de evidências de carga por três meses
Confiabilidade/Credibilidade	Realização de treinamentos e <i>workshops</i> Maior envolvimento por parte da área usuária durante o período de construção, testes e homologação

Tabela 10 - Medidas propostas para resolver os problemas das dimensões da qualidade de dados
Fonte: A autoria do próprio autor.

As medidas apresentadas estão em linha com a teoria levantada no item 2.9 Qualidade dos dados do *Data warehouse*, provendo medidas para incentivar o uso dos dados por parte da área de negócios, envolvendo-os no processo como um todo e diminuindo a sua desconfiança.

O autor do trabalho passou grande parte do projeto atuando como suporte e ponto focal para o processo, ficando alocado juntamente com a área de negócios usuária do sistema. Durante esse período, o autor pode viver na prática a falta de confiança que os usuários da área financeira tinham sobre os dados armazenados, preferindo utilizar bases paralelas. O autor do trabalho também ficou responsável pela elaboração da documentação de evidência de carga dos dados de três meses, comparando os dados contidos nos arquivos de origem com os dados contidos no *data warehouse*. Foram realizados diversos testes e cargas de dados antes da liberação para uso dos usuários, pois se queria garantir a total acurácia dos dados contidos dentro do *data mart*.

3.9 Conclusões

O modelo adotado pela Empresa Beta para o seu sistema de *data warehouse* está alinhado com a abordagem de Kimball, que propõe a carga dos dados diretamente nas bases departamentais conhecidas como *data marts*, em contraposição ao modelo proposto por Inmon de uma base única, a partir da qual são criadas as bases departamentais. Tal escolha foi motivada principalmente pela urgência da primeira aplicação de *data warehouse* e pelo orçamento limitado da Empresa Beta. Outro fator considerado na escolha da abordagem foi o escopo limitado ao qual o sistema deveria atender, contemplando somente o processo de negócios da área financeira.

Outra consideração a respeito da macroestrutura do *data warehouse* estudado diz respeito à adoção da ODS e do seu papel diferenciado dentro do modelo, criando um desenho diferente do encontrado na revisão bibliográfica. A adoção desse modelo diferente é explicada pela necessidade de diminuição do tempo de processamento e carga, que passou de 6 horas para apenas 20 minutos.

Observando o desenvolvimento do projeto para a readequação da solução, percebeu-se que as fontes de dados não sofreram alterações e que as alterações seriam necessárias na camada de modelagem financeira, representada pelo software Cognos Planning. Como consequência das mudanças na modelagem financeira, foi necessário alterar também o modelo do *data warehouse*, nesse caso específico, do *data mart* financeiro. Foram alterados os modelos do tipo *star-schema* das fatos de Custo Despesa, Custo Opex e DRE, de modo a incluir informações relacionadas ao novo formato exigido pelo grupo controlador estrangeiro.

A questão da qualidade de dados no *data warehouse* também foi contemplado na solução, com medidas que buscavam a inclusão dos usuários no processo de construção e de homologação da solução, de modo a incentivar os usuários a utilizar os dados presentes no *data warehouse*. Também foi tratada a questão da necessidade de treinamentos e *workshops* para os usuários da área de negócios, explicando aspectos mais técnicos do processo e como estes poderiam influenciar na qualidade de dados final.

3.10 Resultados

O primeiro e principal resultado obtido após a conclusão do projeto e a readequação do sistema de *data warehouse* para as novas necessidades da Empresa Beta é a aderência do processo à nova formatação dos dados financeiros. Com todo o sistema preparado para receber as informações relativas ao departamento financeiro, agora alinhadas aos requerimentos do grupo controlador estrangeiro, o sistema de *data warehouse* passou a ser mais utilizado, voltando a ser a base de dados oficial da área financeira. A consolidação via Excel e a carga desses dados na base paralela parou de ocorrer depois de decorrido um mês da entrega da solução. A tarefa de consolidação passou a ser realizada via software específico para essa finalidade, o Cognos Planning, que por sua vez carrega os dados no *data warehouse*.

Com a utilização do *data warehouse* como a base oficial, é possível comparar novamente os dados elaborados pela equipe de orçamento com os valores efetivamente realizados consolidados pela equipe de fechamento gerencial. Essa comparação permite às pessoas-chave da Empresa Beta acompanhar o desempenho da organização como um todo, de cada regional, de cada centro de custo, permitindo a tomada de ações para reverter possíveis desvios, em um processo contínuo de análise e consumo das informações.

3.11 Considerações Finais

Como identificado na análise do item 3.8.1.1 Análise da arquitetura, o ambiente de *data warehouse* da Empresa Beta pode ser classificado como aderente à abordagem de Kimball, que propõe a utilização de diversas bases departamentais, os *data marts*, que seriam integrados através da arquitetura *Data Warehouse Bus*, conferindo aos diversos *data marts* o caráter de base integrada de *data warehouse*. Um ponto de observação com relação ao caso estudado está na necessidade de maior coerência e integração entre os *data marts* da Empresa Beta, através da adoção plena da arquitetura *Data Warehouse Bus* e da utilização de dimensões coerentes e conformes. O autor do trabalho não identificou nenhum esforço nesse sentido durante o período em que esteve alocado na Empresa Beta, o que é um fato preocupante, pois sem a plena adoção das dimensões coerentes e conformes, o objetivo de

possuir uma base de dados integrada como o *data warehouse* não se concretiza. A empresa somente trocará a falta de integração de dados dos sistemas transacionais para a falta de integração de dados no ambiente de *data warehouse*.

4. CONCLUSÕES

O trabalho de formatura apresentado teve como objetivos o entendimento sobre o tema *data warehouse*, através da revisão bibliográfica, e a análise de uma aplicação de *data warehouse* em uma empresa de telefonia móvel, através da apresentação de um projeto de melhoria.

O mercado das empresas de telefonia móvel está em constante transformação, o que provoca alterações nas necessidades e nas regras de negócio com o passar do tempo. Tais mudanças de necessidades tornaram uma parte do sistema de *data warehouse* da Empresa Beta obsoleto, com problemas de aderência às necessidade da área financeira.

Com esse cenário de baixa utilização do sistema e de falta de qualidade dos dados contidos no *data warehouse*, foi desenvolvido um projeto de melhoria para a solução, de modo a atender aos requisitos da área de negócios.

Além de apresentar o projeto de melhoria e suas atividades, foi possível realizar um levantamento da solução inicial e elaborar análises a respeito da solução anteriormente implantada. Com isso, foi possível identificar aspectos teóricos da revisão bibliográfica em um estudo de caso real.

O trabalho também proporcionou o aprendizado sobre o tema de *data warehouse* para o autor, que é um assunto importante nos dias atuais, pois o mesmo é responsável por subsidiar o processo de tomada de decisões dentro das organizações. O estudo bibliográfico sobre *data warehousing* também foi benéfico para o desenvolvimento profissional do autor, pois este tema está intimamente relacionado à sua área de atuação profissional.

5. REFERÊNCIAS BIBLIOGRÁFICAS

ALBERT, George. The importance of data warehousing. BusinessLine Internet edition, May 3, 2000. Disponível em: <<http://www.thehindubusinessline.in/2000/05/03/stories/150339m6.htm>>. Acesso em 11 set. 2012.

ALMEIDA Jr., Jorge Rady de. Qualidade de Dados em Data Warehouse. São Paulo: Escola Politécnica da Universidade de São Paulo. Notas de aula para disciplina de pós-graduação do Departamento de Engenharia de Computação e Sistemas Digitais, PCS5000 – Depósito de Dados. 2011.

AUSTIN, Benny. Kimball and Inmon DW Models. 2010. Disponível em: <<https://bennyaustin.wordpress.com/2010/05/02/kimball-and-inmon-dw-models/>>. Acesso em 05 ago. 2012.

BRESLIN, M. Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models. Business Intelligence Journal, v. 9, n. 1, p. 6-20, 2004.

CALVANESE, D. et al. Enterprise modeling and Data Warehousing in Telecom Italia. Information Systems, v. 31, n. 1, p. 1-32, 2006.

DATA WAREHOUSE HELP. Kimball vs Inmon *Data Warehouse* Architecture. Disponível em: <<http://www.datawarehousehelp.com/data-warehouse-architecture/kimball-vs-inmon/>>. Acesso em 05 ago. 2012.

DEVLIN, B. A. *Fathers of the Data Warehouse*. 2009. Disponível em: <http://www.b-eye-network.com/blogs/devlin/archives/2009/08/fathers_of_the_data_warehouse_3.php>. Acesso em 06 ago. 2012.

DEVLIN, B. A.; MURPHY, P. T. An Architecture for a Business and Information System. *IBM Systems Journal*, v. 27, n.1, p. 60-80, 1988.

ECKERSON, W. *Smart Companies in the 21st Century: The Secrets of Creating Business Intelligence Solutions*. 2003. Disponível em: <<http://tdwi.org/research/2003/07/executive-summary-smart-companies-in-the-21st-century-the-secrets-to-creating-successful-business-int.aspx>>. Acesso em 10 out. 2012.

HARRIS, Jim. *The Metadata Crisis*. 2011. Disponível em: <<http://www.information-management.com/blogs/metadata-BI-data-decisions-Gleick-OCDQ-10021418-1.html>>. Acesso em 05 ago. 2012.

HAYES, Frank. *The Story So Far*. Computerworld.com. 2002. Disponível em: <https://www.computerworld.com/s/article/70102/The_Story_So_Far?taxonomyId=009>. Acesso em 07 ago. 2012.

INMON, W. H. *Building the data warehouse*. Nova York: Wiley Computer Pub., 1996. 401p.

INMON, W. H. Who Coined the Term "*Data Warehouse*"? 2009. Disponível em: <<http://www.b-eye-network.com/view/11054>>. Acesso em 06 ago. 2012.

INMON, W. H.; HACKATHORN, R. D. *Using the data warehouse*. Nova York: Wiley, 1994. 285p.

JIANG, B. Is Inmon's *Data Warehouse* Definition Still Accurate?. 2012. Disponível em: <<http://www.b-eye-network.com/view/16066>>. Acesso em 08 ago. 2012.

KIMBALL, R. The *data warehouse* toolkit: the complete guide to dimensional modeling. Nova York: Wiley, 2002. 436p.

KUMAR, Naveen. Inmon versus Kimball. Disponível em: <<http://people.rit.edu/nxk2100/images/Inmon%20versus%20Kimball.pdf>>. Acesso em 05 ago. 2012.

MACHADO, F. N. R. Tecnologia e projeto de *data warehouse* : uma visão multidimensional. São Paulo: Érica, 2006. 318 p.

MATTIODA, R. A.; FAVARETTO, F. Qualidade de dados e produtos – Uma analogia. In: CONGRESSO NACIONAL DE EXCELÊNCIA EM GESTÃO – “SUSTENTABILIDADE NAS ORGANIZAÇÕES BRASILEIRAS: PÚBLICAS, PRIVADAS E DO TERCEIRO SETOR”, 3, Ago. 2006, Niterói.

MATTIODA, R. A.; FAVARETTO, F. Qualidade da informação em duas empresas que utilizam *Data Warehouse* na perspectiva do consumidor de informação: um estudo de caso. Gest. Prod., São Carlos, v. 16, n. 4, Dez 2009.

ORR, Ken. Data Quality and Systems. Communications of the ACM. Nova York, v. 41, n.2, p. 66-71. Fev 1998.

SINGH, H. S. *Data warehouse* – conceitos, tecnologias, implementação e gerenciamento. São Paulo: Makron Books, 2001. 384 p.