



FANESE - FACULDADE DE ADMINISTRAÇÃO E NEGÓCIOS DE SERGIPE
MBA EM ADMINISTRAÇÃO DE BANCO DE DADOS

JESSICA DA SILVA SANTOS

MODELAGEM DE BANCO DE DADOS RELACIONAL E DIMENSIONAL COMO
FERRAMENTAS DE SUPORTE À DECISÃO

Aracaju

2016

JESSICA DA SILVA SANTOS

**MODELAGEM DE BANCO DE DADOS RELACIONAL E DIMENSIONAL COMO
FERRAMENTAS DE SUPORTE À DECISÃO**

Trabalho de Conclusão de Curso apresentado ao Núcleo de Pós-Graduação e Extensão da FANESE, como requisito para obtenção do título MBA em Administração de Banco de dados.

Aracaju

2016

MODELAGEM DE BANCO DE DADOS RELACIONAL E DIMENSIONAL COMO FERRAMENTAS DE SUPORTE À DECISÃO

Jessica Da Silva Santos¹

RESUMO

No cenário atual da área administrativa, as organizações precisam a todo o tempo tomar decisões com base no seu modelo de negócio e situações já vivenciadas, tendo como base os dados de seus sistemas, a forma de acesso à informação e a sua disponibilidade é crucial para o sucesso dos negócios. Este trabalho de pesquisa científica abrange os conceitos de dois modelos, Relacional e Dimensional, para modelagem de banco de dados e relaciona os casos em que cada modelo é mais utilizado. Para analisar a necessidade de consulta eficiente das informações, buscou-se por meio de uma pesquisa qualitativa do tipo bibliográfica demonstrar que o modelo dimensional é o mais indicado para banco de dados com informações históricas, sendo o modelo mais viável para banco de dados com suporte a consultas gerenciais que se enquadram no tema deste artigo: Modelagem de banco de dados Relacional e Dimensional como ferramentas de suporte à decisão.

Palavras chaves:

Modelo Relacional; Modelo Dimensional; Banco de Dados; Suporte à decisão.

¹ Aluna do Curso de MBA em Administração de Banco de Dados

LISTA DE FIGURAS

Figura 1: Cubo de dados representando Vendas	12
Figura 2: Esquema Estrela representando o Fato Vendas	14
Figura 3: <i>Data warehouse</i>	15
Figura 4: Um projeto de Banco de Dados Relacional	16

LISTA DE TABELAS

Tabela 1- Comparativo entre processamento Transacional e Analítico	15
--	----

SUMÁRIO

1	INTRODUÇÃO	6
2	REVISÃO DA LITERATURA.....	7
2.1	Modelo Relacional.....	8
2.2	Modelo Dimensional	10
2.3	<i>Data Warehouse</i>	14
3	ESCOLHA DO MODELO	15
4	METODOLOGIA	17
5	CONCLUSÃO.....	17
6	REFERÊNCIAS.....	18

1 INTRODUÇÃO

O aumento da competitividade fez as organizações se preocuparem em aperfeiçoar seus serviços através do uso de sistemas informatizados, na década de 1960 estes sistemas armazenavam suas informações cada um em um local e formato diferentes, sem nenhum relacionamento entre os diferentes sistemas, e comumente havia divergências entre os seus cadastros.

Para atender esta demanda de integridade dos dados, na década de 1970, Codd sugeriu o modelo Relacional com objetivo de reduzir a redundância dos dados. E, nesta mesma época, surgiu o termo OLTP (*On line Transaction Processing*) – Processamento de Transações *Online*, para definir o processamento efetuado pelos sistemas transacionais ou operacionais (COLAÇO, 2004).

Os sistemas OLTP comumente utilizam a modelagem Relacional para garantir segurança e eficiência na inserção de dados, e reduzir ao máximo a redundância de informações, porém em uma análise gerencial esses sistemas demoram demasiado tempo para retornar as informações devido aos dados estarem separados de forma fragmentada em diferentes tabelas.

Os sistemas OLTP melhoraram a forma de armazenar as informações, mas não foram inicialmente projetados para integrá-las, eram desenvolvidos sob demanda e cada área da organização tinha o seu sistema isolado dos demais, permanecendo o problema de desatualização dos dados.

Nas últimas décadas, o uso dos sistemas informatizados cresceu significativamente e possibilitou o armazenamento histórico de grandes volumes de dados, um estudo realizado por pesquisadores na Carolina do Sul, Estados Unidos, revelou que a quantidade de dados produzida no período compreendido entre 1986 e 2007 se aproxima dos 296 exabytes, o equivalente a 296 bilhões de gigabytes (Bueno; Viana, 2012), as novas tecnologias fazem crescer este número de dados exponencialmente.

Tendo em vista bases de dados volumosas torna-se muito difícil fazer algumas operações que envolvam relacionamentos entre várias tabelas sem um demasiado tempo de espera que compromete o ambiente operacional, pois havendo uma transação de consulta para um relatório as outras transações ficam impossibilitadas de serem salvas antes do término do acesso anterior, ocorrendo

concorrência entre os dados, considerando que o relatório está buscando a informação no mesmo banco de dados do sistema de uso operacional.

Dessa forma, a preocupação em pensar em estruturas de armazenamento com capacidade de acessar facilmente dados históricos de forma transparente para os usuários com menor tempo de resposta possível e flexibilidade para manipulação é o motivador para o presente artigo. Além da melhoria de desempenho, a forma como os dados são estruturados auxiliam no processo de descoberta de conhecimento ao se utilizar técnicas como mineração de dados.

O objetivo deste trabalho é gerar subsídios para a identificação da melhor forma de modelagem de dados para descoberta de conhecimento em bancos de dados com grande quantidade de informações que auxiliem a tomada de decisões demonstrando os Modelos Relacional e Dimensional para banco de dados. Será realizada uma revisão a bibliografia das modelagens de dados, comentando qual melhor modelo para diferentes situações, e feita uma abordagem dos conceitos da arquitetura *Data Warehouse*, que utiliza a Modelagem Dimensional.

Este trabalho se justifica na necessidade de melhoria no desempenho de consultas de grande quantidade de dados, visto que as organizações cada vez mais informatizam seus processos e tendem a analisar historicamente esses dados para melhoria de competitividade nos negócios ao descobrir, por exemplo, padrões que auxiliem na tomada de decisão. Observando que os sistemas do ambiente operacional utilizam comumente o modelo relacional, para gerar relatórios relacionando diferentes bancos de dados desses sistemas o processo é complexo e a geração dos relatórios é demorada, a modelagem dimensional é a melhor solução para melhor desempenho das consultas de grande volume de dados armazenados historicamente e para flexibilidade na construção de relatórios gerenciais.

2 REVISÃO DA LITERATURA

De acordo com ELMASRI (2011) um modelo de dados é uma coleção de conceitos que podem ser usados para descrever a estrutura de um banco de dados que são os tipos, relacionamentos e restrições que se aplicam aos dados.

A elaboração do modelo de dados concentra-se na observação dos fatos relevantes que ocorrem na realidade, com a finalidade de construir um sistema que

possa automatizar as necessidades de informação da mesma. Segundo MACHADO (1996), os documentos que registram estes fatos só devem ser utilizados como apoio de entendimento, e não como base para o desenvolvimento do sistema de informações, ou seja, não se deve ter a preocupação em simular o ambiente atual, seja ele manual ou automatizado.

Os modelos de dados são classificados de acordo com a arquitetura que utilizam, neste trabalho será apresentado o Modelo Relacional, surgido para atender as demandas de sistemas transacionais, OLTP, e o Modelo Dimensional, que atende aos sistemas analíticos, OLAP.

2.1 Modelo Relacional

No início da popularidade dos sistemas informatizados vários sistemas eram desenvolvidos nos departamentos das empresas, cada um com sua base separada e um cadastro de clientes e produtos, por exemplo, podia estar armazenado em diversos locais causando muitas vezes inconsistências em relatórios extraídos dos diferentes sistemas se estivessem desatualizados um com relação aos outros, além disso, era difícil realizar uma consulta vinculando dados dessas bases distintas que muitas vezes estavam armazenadas em formatos diferentes.

Os fundamentos de bancos de dados relacionais surgiram na empresa IBM, nas décadas de 1960 e 1970, através de pesquisas de funções de automação de escritório. Neste período as empresas já sinalizavam a necessidade de armazenar seus dados de forma estruturada impulsionando pesquisas que desenvolveram muitas tecnologias usadas nos dias atuais. Em 1970 Ted Codd, pesquisador da IBM, publicou o primeiro artigo sobre bancos de dados relacionais o qual tratava sobre o uso de cálculo e álgebra relacional para permitir que usuários não técnicos armazenassem e recuperassem grande quantidade de informações (SANCHES, 2012).

O Modelo Relacional surgiu para atender sistemas transacionais, utilizados nos processos do dia a dia, com operações atômicas, geralmente com um grande número de usuários simultâneos realizando operações repetidamente que devem ocorrer por completo ou então serem desfeitas, tendo como conceito base separar os dados em tabelas relacionadas em forma de relações matemáticas.

Para aprimoramento do Modelo Relacional existe o processo denominado normalização, neste processo é possível, gradativamente, substituir um conjunto de entidades e relacionamentos por um outro, o qual se apresenta consistente em relação às anomalias de atualização (inclusão, alteração e exclusão), as quais podem causar problemas, tais como: grupos repetitivos (atributos multivalorados) de dados, redundância de dados desnecessária, perdas acidentais de informação, dificuldade na representação de fatos da realidade observada (MACHADO, 1996).

Os bancos de dados que utilizam esse modelo suportam processamento de grande número de transações num tempo curto para processamento, características de sistemas do ambiente OLTP, operacional.

Um projeto OLTP totalmente normalizado para um sistema de controle de pedidos, por exemplo, pode envolver dezenas de tabelas e tornar muito difícil o processo de obtenção de informações a partir destas tabelas. Isso porque o Modelo Relacional foi desenvolvido para atender aos sistemas operacionais e a normalização evita redundâncias de tal forma que não haja preocupação com o sincronismo de dados nas operações de atualização, facilitando a manutenção da integridade desses dados nas tabelas. (UNICAMP, 2006)

O ambiente OLTP é projetado para uma grande quantidade de consultas e gravação de dados sendo, em geral, pouco flexíveis para criação de relatórios gerenciais. O acesso aos dados neste ambiente é rápido e organizado, melhorando o desempenho dos processos operacionais, mas não permite maior detalhamento de dados do que ele já apresenta.

A técnica de Modelagem de dados Relacional mais difundida e utilizada é a abordagem Entidade Relacionamento (ER). Nesta técnica, o modelo de dados é representado através de um modelo Entidade Relacionamento. Geralmente, um modelo Entidade Relacionamento é representado graficamente através de um Diagrama Entidade-Relacionamento (DER). Esta abordagem foi criada em 1976 por Peter Chen, podendo ser considerada como um padrão de fato para a modelagem conceitual (BRASIL, 2012).

Kimball, 1997, define o Modelo Entidade Relacionamento como uma técnica de projeto lógico que busca remover a redundância de dados. E, sendo relacional, é imensamente benéfica aos processos transacionais porque as transações são feitas de forma determinada e simples. No entanto, os processos de negócio, nessa

modelagem, são vistos de forma separada e relacionar os dados pode se tornar uma tarefa confusa.

O Modelo Entidade Relacionamento para uma empresa que possui centenas de entidades lógicas possui, no geral, para cada uma destas entidades uma tabela física, quando o banco de dados é implementado torna-se complexo entender os relacionamentos dos dados entre as tabelas e realizar uma consulta de análise estratégica.

2.2 Modelo Dimensional

Os sistemas operacionais ocasionaram o armazenamento de grande quantidade de dados históricos estruturados, disponíveis para serem utilizados no auxílio dos processos de tomada de decisão, porém fragmentados em diferentes bases. O Modelo Dimensional surgiu para atender esta demanda de processamento analítico, com consultas para análise estratégica do negócio.

Segundo KIMBALL (1997), a Modelagem Dimensional é uma técnica de projeto lógico que procura apresentar os dados com um padrão intuitivo, construída para permitir alto desempenho em ambientes com grande quantidade de dados armazenados historicamente. Todo Modelo Dimensional é composto por uma tabela multi chaves, chamada fato e um conjunto de tabelas chamadas dimensão. Cada tabela tem uma única chave primária que corresponde a um dos componentes da multi chave da tabela fato, a representação visual destas tabelas em que a tabela fato fica situada no centro e as dimensões dispostas ao redor, formam um desenho semelhante a uma estrela, é o que denomina usualmente essa estrutura.

A tabela de fatos é derivada da combinação de chaves que definem cada registro e também contêm uma ou mais medidas numéricas. A adição dessas medidas é importante para recuperar registros em uma tabela única quando se recuperam uma enorme quantidade de registros. Os dados nessas tabelas são desnormalizados para garantir melhor desempenho evitando criar muitos relacionamentos entre várias tabelas como no Modelo Relacional.

As tabelas de dimensão, na maioria das vezes, contêm informações textuais descritivas. O Modelo de banco de dados Dimensional ao ser apresentado para os usuários finais é altamente reconhecível por eles como reflexo da sua estrutura de

negócio. Outra característica desse Modelo é que permite a navegação de um nível de detalhamento para o outro nível “mais baixo” ou “mais alto” em tempo real.

Como explicado em TECHNET (2012), as dimensões permitem filtrar, agrupar e rotular os dados. Por exemplo, você pode filtrar e agrupar pessoas de uma base de dados em categorias por sexo ou idade. Em seguida, os dados podem ser apresentados em um formato no qual são classificados naturalmente pelas hierarquias e categorias, para permitir uma análise mais aprofundada. As dimensões também possuem hierarquias naturais para permitir que os usuários façam “*drill down*” em níveis mais detalhados. Por exemplo, a dimensão Data possui uma hierarquia que pode ser detalhada sucessivamente por ano, semestre, trimestre, mês, semana e dia.

Os bancos de dados dimensionais surgiram para melhorar a relação entre custo e benefício nos ambientes OLAP, de processamento analítico.

Grandes volumes de dados, dados históricos e bases não normalizadas são algumas das peculiaridades que impedem a utilização de ferramentas convencionais para criação de relatórios. Ao deparar-se com esse quadro, a indústria de software, aliada a pesquisadores da área, também passou a investir na concepção de um paradigma de ferramenta que pudesse atender a essa demanda. Desse trabalho, surgiu o que chamamos OLAP (*Analytic Processing On-Line* ou processamento analítico *on-line*). O termo foi criado em 1993, pelo Dr. E.F. (Ted) Codd. (COLAÇO, 2004)

Uma aplicação OLAP tem a capacidade de visualização das informações a partir de várias perspectivas diferentes mantendo uma estrutura de dados eficiente e tem por finalidade apoiar os usuários finais a tomar decisões estratégicas.

Normalmente, o sistema OLAP é projetado para atender um pequeno número de usuários que realizam consultas planejadas com um tempo de resposta maior, devido a consultas operarem sobre um grande volume de dados, há redundância planejada dos dados, é geralmente implementado para ambientes de *Data Warehouse* onde compensa pela menor quantidade de atualização das informações. Nesse sistema, o usuário realiza apenas consultas na base de dados que é estruturada para armazenar dados históricos, tipicamente de 12 a 60 meses, conforme Hokama (2004).

Os sistemas OLAP fornecem uma visão Multidimensional dos dados independentemente de como eles estão fisicamente armazenados. Os dados são

percebidos para o usuário como uma estrutura conhecida como Cubo Multidimensional.

Um array Multidimensional tem um número fixo de dimensões e os valores são armazenados nas células. Cada dimensão consiste de um número de elementos. Em outras palavras, o Cubo é, de fato, apenas uma metáfora visual, uma representação intuitiva do evento, pois todas as dimensões coexistem para todo ponto no Cubo e são independentes umas das outras (COLAÇO, 2004).

O Cubo expressa a forma na qual os tipos de informações se relacionam entre si, de forma genérica, armazena todas as informações sobre um determinado assunto, fato, de forma que permite várias combinações entre elas em vários níveis de detalhamento.

No artigo da Microsoft (TECHNET, 2012) é explicado que os Cubos podem exibir e somar grandes volumes de dados, embora também forneçam aos usuários acesso pesquisável a quaisquer pontos de dados para que os dados possam ser acumulados, decompostos e analisados, conforme a necessidade a fim de tratar da maior variedade de questões relevantes à área de interesse do usuário.

A ilustração a seguir mostra um Cubo OLAP que contém as dimensões de Tempo, Região e Produto.



Fonte: TECHNET (2012).

Data: 2012.

Dentro de cada dimensão de um modelo OLAP, os dados podem ser organizados em uma hierarquia que define diferentes níveis de detalhe. A dimensão tempo, por exemplo, tem uma hierarquia representando os níveis ano, mês e dia. Da mesma forma, a dimensão região tem os níveis país, região, estado e cidade. Assim, um usuário visualizando dados em um modelo OLAP irá visualizar os níveis acima (*drill up*) ou abaixo (*drill down*) navegando entre os níveis para visualizar a informação com maior ou menor nível de detalhe.

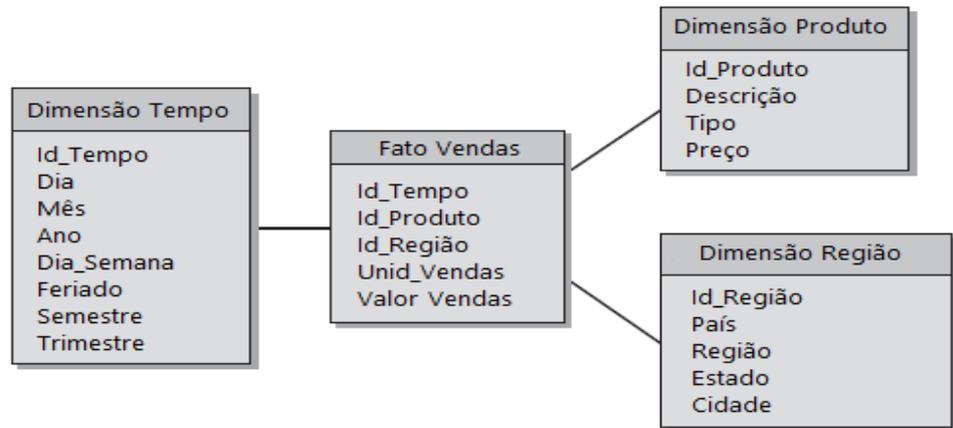
Como exemplo de consulta, um gerente de negócios pode conseguir um resumo rápido e simples das vendas do produto 1 no ano 2015. Esse mesmo gerente pode detalhar ainda mais para obter números de vendas para um período de tempo mais direcionado.

Outro agrupamento pode trazer dados para examinar como as vendas foram afetadas por determinado acontecimento no decorrer do tempo. Isso ajuda a determinar quais tendências de vendas estão ocorrendo e quais são as possíveis revisões necessárias à estratégia de negócios. Por meio da filtragem na dimensão tempo, essas informações podem ser rapidamente fornecidas e consumidas. Essa segmentação e análise de dados é possível porque as dimensões no modelo dimensional são projetadas com atributos e dados que podem ser facilmente filtrados e agrupados pelo cliente (TECHNET,2012).

Outra forma de representar a modelagem dimensional é através de uma estrutura denominada *Star Schema*, traduzindo Esquema Estrela.

O conceito de Esquema Estrela foi criado por Ralph Kimball e tem como característica básica a presença de dados altamente redundantes para se obter melhor desempenho (COLAÇO, 2004). O nome Esquema Estrela é devido a disposição das tabelas de forma semelhante a uma estrela. Este esquema é composto por uma tabela central de fatos rodeada de tabelas dimensões. A figura 2 exemplifica uma estrutura simples do modelo representando o cubo de vendas da figura 1.

Figura 2:
ESQUEMA ESTRELA REPRESENTANDO O FATO VENDAS



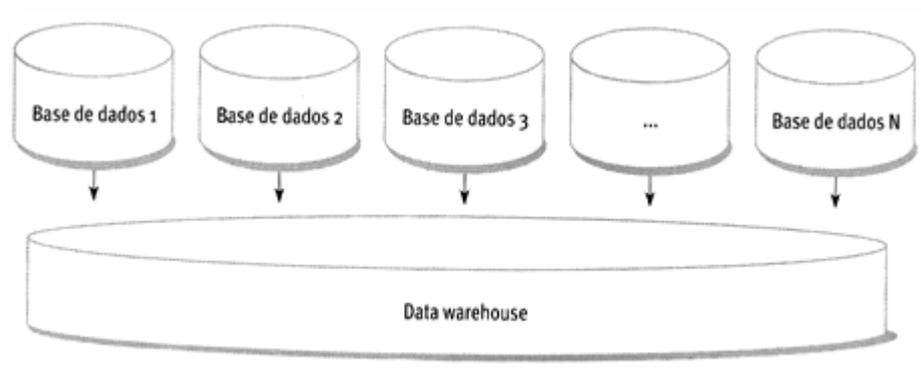
Fonte: Adaptado de COLAÇO (2004).

Data: 2004, p. 50.

2.3 Data Warehouse

Um modelo de dados bem estruturado é capaz de prover à empresa a capacidade de extrair informações das mais diversas maneiras (HOKAMA, 2004). Quando se tem os dados distribuídos em várias bases de dados, muitas vezes é necessário agrupar as informações em um banco organizado para o suporte a tomadas de decisões estratégicas. O conceito de *Data Warehouse* surgiu da necessidade de integrar dados de diferentes bases em um esquema unificado como ilustra a figura 3.

Figura 3
DATAWAREHOUSE



Fonte: CARDOSO; CARDOSO (2013).

Data: 2013, p.188.

A Modelagem Dimensional é a técnica de modelagem normalmente utilizada para implantação de um *Data Warehouse*, sendo voltada especificamente para essa finalidade.

Uma consideração importante citada por COLAÇO (2004), é que o tempo de resposta das consultas realizadas sobre o *Data Warehouse* dependerá não somente dos recursos de rede, dependerá também do volume de informações acessadas. Dependendo da pesquisa, o tempo de resposta poderá ser de alguns minutos independentemente da velocidade dos recursos de rede.

3 ESCOLHA DO MODELO

Os ambientes transacionais e analíticos possuem características diferentes, logo necessidades também distintas. Para cada um desses ambientes um modelo é mais adequado, a Tabela 1, abaixo, resume as diferenças entre os dois ambientes.

Tabela 1:

COMPARATIVO ENTRE PROCESSAMENTO TRANSACIONAL E ANALÍTICO

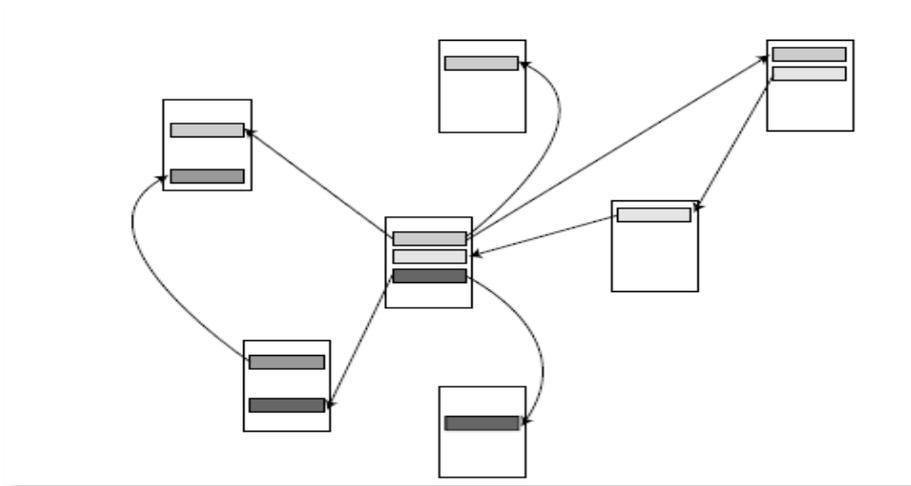
Características	Sistemas Transacionais (OLTP)	Sistemas analíticos (OLAP)
Atualizações	Mais frequentes	Menos frequentes
Tipo de informação	Detalhes	Agrupamento
Quantidade de dados	Poucos	Muitos
Precisão	Dados atuais	Dados históricos
Complexidade	Baixa	Alta
Consistência	Microscópica	Global
Exemplos	CRM, ERP, Supply Chain	MIS, DSS, EIS
Terminologia	Linhas e Colunas	Dimensões, Medidas e Fatos

Fonte: Adaptado de CUNHA,

Data:2015.

A abordagem Relacional para banco de dados organiza os dados em uma tabela e utiliza chaves primárias e chaves estrangeiras para relacioná-los. A Figura 4 exibe um design clássico de banco de dados relacional, demonstrando que o relacionamento entre os dados está em várias tabelas sucessivas, e para realizar consultas de informações nesse modelo é necessário um grande número de junções o que torna maior o tempo de resposta.

Figura 4:
UM PROJETO DE BANCO DE DADOS RELACIONAL



Fonte: INMON (2005).

Data: 2005, p.358.

A estrutura Multidimensional possui desnormalização de tabelas, alta redundância e suporta periodicidade de atualizações de dados muito menor do que uma estrutura Relacional.

KIMBALL (1997) explica que é possível converter um modelo Relacional para Dimensional de forma que um diagrama Entidade Relacionamento se divide em vários diagramas de Dimensões, pois nele é representado cada processo de negócio de uma empresa.

Para converter um diagrama Relacional em um diagrama Dimensional é necessário relacionar os processos e selecionar as relações existentes, definir a tabela de fatos e as tabelas que se conectam diretamente a esta, tabelas de dimensão.

O Modelo Dimensional resulta uma interface mais compreensível ao usuário final e proporciona um maior desempenho nas consultas de grandes quantidades de dados sendo a única técnica viável para bancos de dados que são projetados para suportar consultas de usuários finais em um *Data Warehouse*.

4 METODOLOGIA

A metodologia utilizada no artigo foi qualitativa do tipo bibliográfica com os seguintes passos: Estudo do conteúdo bibliográfico e análise das duas abordagens: Relacional, Dimensional e a definição do melhor modelo para diferentes situações.

5 CONCLUSÃO

No contexto apresentado neste trabalho, de ferramentas para suporte à decisão, os diagramas Entidade Relacionamento tradicionais, Modelagem Relacional, não são os mais adequados às necessidades de análise dinâmica dos dados, como por exemplo: execução de consultas com desempenho aceitável, tratamento de grandes volumes de dados, suporte a modelos flexíveis de consulta dos dados.

No entanto, os processos de negócio, nessa modelagem, são vistos de forma separada e relacionar os dados se tornou uma tarefa confusa, se a empresa possui um grande número de entidades lógicas cada uma destas geralmente se transforma em uma tabela física, quando o banco de dados é implementado os usuários finais não conseguem entender facilmente o negócio visualizando este modelo.

A modelagem Entidade Relacionamento é direcionada aos dados e não às regras de negócio, logo a melhor Modelagem para um sistema de apoio à decisão que permita identificação de padrões é a Modelagem Dimensional.

A estrutura Relacional diferencia-se da estrutura Dimensional principalmente devido a normalização, pouca redundância e a frequência de atualizações suportadas sendo muito útil para o ambiente transacional, mas deve ser evitado para entrega ao usuário final que necessita de consultas estratégicas para tomada de decisão.

ABSTRACT

In the current scenario of the administrative area, organizations need all the time to make decisions based on its business model and has experienced situations, based on the data from their systems, the way of access to information and availability is crucial to business success. This scientific research work covering the concepts of two models, relational, and dimensional to database modeling and relates to cases in which each template is most often used. To examine the need for effective consultation of information was sought through a qualitative research of bibliographical demonstrate that the dimensional model is the most suitable for database with historical information, the most viable model for database with support the managerial appointments that fall into the topic of this article: Modeling for decision support databases.

Key words:

Relational Model; Dimensional Model; Database; Decision support.

6 REFERÊNCIAS

BRASIL. **Administrador de Banco de dados**. IFIS Paraná - Pronatec, Educação Inicial e Continuada. 2012.

BUENO, M. F.; VIANA, M. R. **Mineração de Dados: Aplicações, Eficiência e Usabilidade**. In: Congresso de Iniciação Científica do INATEL (INCITEL), 2012. Anais. Santa Rita do Sapucaí (MG): Instituto Nacional de Telecomunicações (INATEL), 2012.

COLAÇO JÚNIOR, M. **Projetando Sistemas de Apoio à Decisão Baseados em Data Warehouse**. Rio de Janeiro: Axcel Books, 2004.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados**. 6ª edição. São Paulo. 2011.

GALVAO, N. D.; MARIN, H. F. **Técnica de mineração de dados**: Uma revisão da literatura. São Paulo. Acta paul. enferm. vol.22 no.5, 2009.

HOKAMA, D.D.B.; CAMARGO, D.; FUJITA, F.; FOGLIENE, J.L.V. **A Modelagem de Dados no Ambiente Data Warehouse**. Universidade Presbiteriana Mackenzie – Faculdade de Computação e Informática. São Paulo. 2004.

INMON, W.H. **Building the Data Warehouse**. Fourth Edition. Wiley. 2005.

KIMBALL, R. **A Dimensional Modeling Manifesto**. DBMS and Internet Sysytems. 1997.

_____. R. **The data warehouse ETL toolkit** : Practical techniques for extracting, cleaning, conforming, and delivering data. Wiley, 2004.

LAROSE, D. T. **Discovering Knowledge in Data**: An Introduction to Data Mining. John Wiley and Sons, Inc, 2005.

MACHADO, M. P. A. **Projeto de Banco de Dados**: Uma visão prática. São Paulo. Érica, 1996.

SANCHES, A. R. **Fundamentos de Armazenamento e Manipulação de Dados**: Introdução e história. Disponível em: <http://www.ime.usp.br/~andrrs/aulas/bd2005-1/aula2>. Acesso em 15 mai. 2016.

UNICAMP, Revista. **Informativo Técnico no. 54**: Análise Multidimensional. Disponível em: < <http://www.ccuec.unicamp.br/revista/infotec/informacao/inf54.html>>. Acesso em 07 mai. 2016.

TECHNET. **Noções básicas sobre cubos OLAP**. 2012. Disponível em: < [https://technet.microsoft.com/pt-br/library/hh916543\(v=sc.12\).aspx](https://technet.microsoft.com/pt-br/library/hh916543(v=sc.12).aspx)>. Acesso em 19 mai. 2016.