

## DATA WAREHOUSE A VANTAGEM DA MODELAGEM DIMENSIONAL DE DADOS

SILVA, Marcio.<sup>1</sup>  
SARTORI, Marco.<sup>2</sup>

### RESUMO

O trabalho teve como objetivo verificar a vantagem da aplicação do modelo dimensional de dados no que se refere a consultas direcionadas a tomadas de decisões no ambiente corporativo. Um dos problemas enfrentados pelas empresas é a falta de agilidade e qualidade das informações que são utilizadas para decisões estratégicas. Isso se deve ao fato que, normalmente, os dados disponíveis não estejam devidamente organizados para oferecer tal agilidade e qualidade. Essa necessidade foi observada na empresa onde foi realizado o estudo de campo, pois possui uma grande base de dados e sistemas que atendem bem o ambiente operacional, mas não oferece resultados satisfatórios quando se trata de tomada de decisão. Atualmente, é de extrema importância que as empresas possuam respostas rápidas e confiáveis que possibilitem ações acertadas e em tempo hábil em um mercado altamente competitivo.

**PALAVRAS-CHAVE:** Banco de dados, Data warehouse, Business Intelligence, Modelagem dimensional, Tomada de decisão

## DATA WAREHOUSE THE ADVANTAGE OF DIMENSIONAL DATA MODELING

### ABSTRACT

This article aimed to verify the advantage of applying dimensional data model with regard to queries directed to decision-making on the corporate environment. One of the problems faced by companies is the lack of speed and quality of information that is used for strategic decisions. This is due to the fact that, usually, the available data are not organized to provide such speed and quality. This necessity was seen in the company where the study was carried out because it has a large database and system that serve well the operating environment, but does not provide satisfactory results when it comes to decision-making. Currently, it is extremely important that businesses have reliable and fast answers that will provide right actions in real time on today's highly competitive market.

**KEY WORDS:** Data bank, Data warehouse, Business Intelligence, Dimensional modeling, decision making.

## 1. INTRODUÇÃO

Atualmente, com o mundo globalizado e cada vez mais competitivo, os avanços na tecnologia da informação proporcionam aos negócios ferramentas poderosas, que possibilitam às corporações a capacidade de cruzar dados de maneiras diversas e de forma extremamente rápida e segura. Devido à alta complexidade envolvendo a tomada de decisão, é fundamental a análise de dados que permitam respostas mais acertadas e em tempo hábil. As empresas possuem grandes volumes de dados operacionais, voltados ao controle das operações diárias, mas, não é simples utilizar esses dados para a tomada de decisões, pois sua organização pode ser complexa e demorada.

*Data warehouse* vem ao encontro dessas necessidades, pois, através de técnicas específicas, permite que dados já existentes na corporação, independente de qual seja a sua fonte, sejam organizados e dispostos de forma a proporcionar uma maneira diferente de tratá-los, o que abre um novo e promissor horizonte de possibilidades.

Este artigo é relevante ao analisar as vantagens da utilização de *data warehouse* no ambiente computacional corporativo, como forma de melhorar o desempenho das consultas de dados voltadas à tomada de decisão e com isso também diminuir a demanda de consultas no banco de dados operacional em busca da resposta à pergunta: Como tratar os dados de fontes diversas, existentes no ambiente corporativo, transformando-os em informações estratégicas para as tomadas de decisão com precisão e rapidez?

Apesar de não se tratar de uma tecnologia muito recente, muitas empresas ainda não utilizam *data warehouse*, o que gera sobrecarga da sua base operacional de dados com consultas direcionadas a tomadas de decisão que não apresentam a performance esperada e necessária, comprometendo a capacidade da empresa em levantar e analisar informações de forma rápida e segura, impedindo-a de agir com a rapidez exigida em um cenário altamente competitivo. Atualmente, as empresas possuem bases de dados muito grandes e complexas, onde não é tão simples o trabalho de análise desses dados. De acordo com Singh (2001), nossa capacidade de colher e armazenar dados já superou nossa habilidade de analisar e transformar esses dados em conhecimento. Ainda segundo Singh (2001), as organizações possuem dados em abundância, redundantes e inconsistentes, difíceis de administrar e cada vez mais difíceis de serem utilizados para fins de suporte à decisão.

<sup>1</sup>Acadêmico – Faculdade Assis Gurgacz. E-mail: marcio.infocus@gmail.com

<sup>2</sup>Docente orientador – Faculdade Assis Gurgacz. Curso de Sistemas de Informação. E-mail:

Diante de todas essas questões, o propósito do trabalho atual é apresentar as vantagens que a modelagem dimensional de dados apresenta em relação ao modelo relacional.

## 2. REFERENCIAL TEÓRICO OU FUNDAMENTAÇÃO TEÓRICA

Para Milani (2006) existem basicamente dois tipos de bancos de dados, sendo um mais antigo, que é baseado em arquivo texto e de acordo com as demandas e a própria evolução tecnológica, foi desenvolvido o modelo no qual é incorporado um sistema gerenciador de banco de dados. De acordo com Heuser (1998), no primeiro modelo, todas as funcionalidades desejadas em um software eram implementadas diretamente na aplicação, incluindo todo o tratamento de controle de persistência dos dados e operações de segurança. Além disso, os cálculos eram escritos diretamente na linguagem de programação utilizada para o desenvolvimento da aplicação. Em um segundo momento, com a necessidade de unir funcionalidades comuns a muitos programas, foram criados os Sistemas Gerenciadores de Banco de Dados (SGBDs). Dessa forma, o tratamento da persistência dos dados, o controle de acesso e diversas outras funções puderam ser transferidas para o banco de dados, tornando menos complicado o trabalho do programador, diminuindo o índice de manutenção e fazendo com que o programa ficasse mais compreensível.

“O grande desafio das organizações é o gerenciamento de dados, sua organização e tratamento para que possibilite uma utilidade muito mais ampla do que a permitida pelos sistemas de banco de dados relacionais”. (RASLAN; CALAZANS, 2014).

As corporações, normalmente, possuem grande volume de dados, que se encontram em sistemas distintos, com dados duplicados, em estruturas completamente diferentes e bases distintas.

Grandes empresas possuem diversas fontes de informação que precisam utilizar para tomar decisões comerciais. As fontes podem armazenar dados sob diferentes esquemas. Por razões de desempenho (também por razões de controle organizacional) as fontes de informação não permitem que outras partes da empresa recuperem dados de acordo com a demanda. (SILBERSCHATZ, 2004, p.703).

Em um ambiente com dados organizados de forma tão complexa, é impossível obter informações em tempo hábil e com confiabilidade para a tomada de decisões. Simples relatórios e consultas em sistemas distintos não trazem a resposta satisfatória para as necessidades corporativas.

### 2.1 DATA WAREHOUSE

Um *data warehouse* tem o objetivo de reunir de forma organizada, dados de fontes diversas, possibilitando que as consultas com objetivo de tomada de decisão sejam feitas, com performance melhorada, em um único banco de dados desenvolvido e otimizado especificamente para tal tarefa. De acordo com Colaço Júnior (2004, p. 16), “*Data Warehouse* (DW) é um banco de dados histórico, separado lógica e fisicamente do ambiente de produção das organizações, concebido para armazenar dados extraídos deste ambiente”.

Os bancos de dados tradicionais possuem informações pobres para a tomada de decisões estratégicas e de risco, a resposta da Tecnologia para resolver este problema foi o Data Warehouse (DW), uma tecnologia que permite aos executivos e usuários experientes tomarem decisões seguras com o menor risco possível, sem a necessidade de mesclar vários relatórios gerenciais e sintetizá-los para executar uma ação. (SASSI, 2000, p. 63).

Segundo Date (2004), os *data warehouses* surgiram pela necessidade de existir uma origem única de dados, limpa e consistente para apoio à decisão e também para que isso seja feito sem causar impacto sobre os sistemas operacionais.

Ainda segundo Colaço Júnior (2004, p. 16), “Antes de serem armazenados no DW, os dados são selecionados, integrados e organizados para que possam ser acessados da forma mais eficiente, auxiliando assim o processo de tomada de decisão”.

#### 2.1.1 Características de um *data warehouse*

Um *data warehouse* possui algumas características que o definem. É orientado por temas, integrado, variante no tempo e não volátil.

De acordo com Colaço Júnior (2004), um *data warehouse* é orientado por temas, pois organiza as informações de acordo com temas importantes para o negócio da empresa, integrado por consolidar dados de diversas origens, variante no tempo, uma vez que os dados históricos são mantidos, e não volátil, uma vez que teoricamente, os dados não poderão ser atualizados ou alterados, apenas acessados. Assim, pode-se definir que um Data Warehouse é:

- Orientado por tema: Um DW está orientado em torno de assuntos específicos da organização.
- Integrado: A própria organização de um DW exige a integração dos dados que nele estão contidos para que

os objetivos sejam alcançados. Essa integração é obtida com o processo de ETL (*Extract Transform and Load*).

- c. Não volátil: Em teoria, os dados são inseridos em um DW, mas nunca excluídos ou alterados. O DW já é preparado para otimizar os processos de inserção e seleção. Mas há casos específicos que exigem a atualização do DW.
- d. Variante no tempo: Um DW apresenta o que se chama de fotos da base operacional no momento, portanto os dados podem ser comparados através do tempo.

Segundo Bonel (2015, p. 56), “O Data Warehouse funciona como uma hierarquia de assunto e seu principal objetivo é organizar a informação de forma multidimensional, para facilitar e otimizar a consulta”.

2.1.2 ETL

Um *data warehouse* deve organizar dados de fontes diversas, assim, é necessário que esses dados sejam lidos, tratados e apropriadamente inseridos no DW. Esse processo é realizado através de ferramentas *Extract, Transformation and Load* (ETL) (Extração, Transformação e Carga). Colaço Júnior (2004) afirma que o processo de ETL consome cerca de 70% do tempo de construção de um DW.

O processo de ETL, sem dúvida é bastante crítico. Por lidar com fontes diversas de dados e necessitar organizá-las, é bastante comum a existência de muitos dados duplicados, mas com estruturas diferentes dependendo de cada fonte. Colaço Júnior (2004, p. 16) afirma que “Os dados devem ser perfeitamente integrados para que ao serem armazenados assumam uma única convenção”.

Sobre DW, Colaço Junior (2004, p. 19) afirma que “[...] normalmente é desenvolvido especificamente para cada empresa, devido à diversidade existente em termos de estrutura de dados [...]” (COLAÇO JUNIOR, 2004, p. 19).

O fluxo de dados começa nas aplicações fontes, e passa por uma área intermediária de armazenamento chamada de *Staging Area* (Área de Estágio). Na *Staging Area* os dados sofrem integração, limpeza e depois são exportados para o DW. A integração consiste na consolidação dos dados de diversas origens, o que geralmente envolve diferentes codificações. (COLAÇO JÚNIOR, 2004, p. 19).

De acordo com Baltzan;Phillips (2012, p.161), para que a qualidade da informação seja mantida, deve-se fazer o processo de limpeza de informação, que é um processo responsável pela seleção, conserto, e descarte de informações inconsistentes ou incompletas.

O processo de extração é realizado com o objetivo de obter os dados das diversas fontes e sistemas operacionais e colocá-los na *staging area*, que é um repositório intermediário, onde ocorre a transformação. A transformação tem como objetivo a preparação dos dados, que pode ser feita através de limpeza, padronização, combinação, sumarização, ou qualquer outra operação que traga a qualidade desejada para a finalidade do DW, através de dados bem elaborados e confiáveis. A carga é o processo que transfere os dados devidamente transformados para dentro do DW. É possível verificar, na Figura 1, os elementos básicos de um *data warehouse* de forma bastante simples.

Figura 1 - Elementos básicos de um *Data Warehouse*



Fonte: Kimball (2002)

2.1.2.1 Metadados

Segundo Colaço Júnior (2004, p.68), “Os metadados são definidos como dados sobre dados, porém a complexidade desses dados no *Data Warehouse* aumenta muito”.

Para Ferreira (2002), metadados trazem informações sobre a estrutura de dados e suas relações dentro do banco de dados, e sua disponibilização permite que usuários possam avaliar a compatibilidade dos dados com as aplicações

2.1.3 A modelagem dimensional

De acordo com Antonio (2006), nos bancos de dados relacionais, os dados são distribuídos em inúmeras tabelas que se relacionam entre si por campos em comum. O objetivo do modelo relacional é possibilitar que operações rotineiras possam ser persistidas de forma rápida e integrada. Essas operações correspondem a grandes volumes de dados, uma vez que são mantidos dados de cada operação de maneira única. Já um *data warehouse*, utiliza o conceito de modelagem dimensional, que é a forma de organização de dados que visa atender as necessidades de consultas para a tomada de decisões. No modelo dimensional, a modelagem ocorre de forma diferente do modelo relacional, pois o foco está nos resultados rápidos exigidos para a tomada de decisões. O quadro 1 apresenta a comparação entre algumas características do modelo dimensional e relacional.

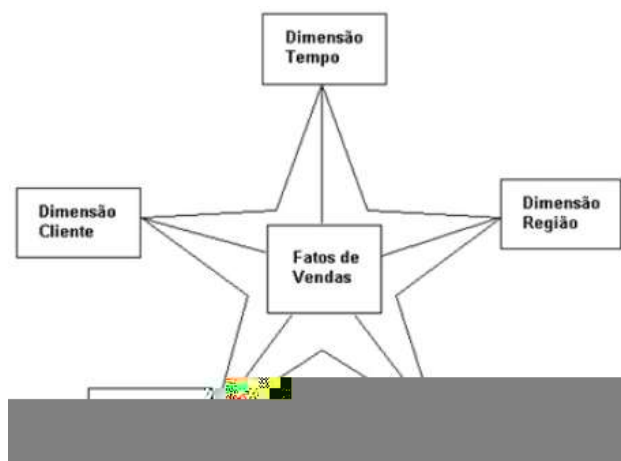
Quadro 1 - Comparação entre Modelo Dimensional e Modelo ER

Modelo Dimensional	Modelo Relacional - ER
Padrão de estrutura mais fácil e intuitiva	Modelo mais complexo
Anterior ao ER, anos 60	Ênfase nos bancos de dados relacionais, anos 70
Tabelas Fato e tabelas dimensão	Tabelas que representam Dados e Relacionamentos
Tabelas Fato são o núcleo – normalizadas	Todas as tabelas são comumente normalizadas
Modelo mais facilmente "joined"	Maior dificuldade de "join", por ter um número maior de tabelas
Leitura mais fácil do modelo por usuários não especializados	Maior dificuldade de leitura pelo usuário não especializado.

Fonte: (BARBIERI, 2001)

Ao considerar algumas relações existentes a uma venda, encontraremos cliente, data da venda, funcionário que efetuou a venda. No esquema conhecido como *Star Schema* (Esquema Estrela), define-se uma tabela fato, a tabela de vendas, e as demais tabelas relacionadas são as tabelas dimensão. No modelo gráfico representando as ligações é possível perceber a semelhança a uma estrela.

Figura 2 - O Esquema estrela



Fonte: Machado (2004)





## 2.2 Data Mart

“Os *data marts* representam um subconjunto dos DW que permitem o acesso descentralizado à informação. Os DM podem ser direcionados a um departamento ou área específica do negócio” (MACHADO, 2006). Entende-se com isso, que *data mart* é um DW de dimensões reduzidas, com custo menor, e tem como objetivo atender apenas uma área do negócio, pois mantém informações sobre um assunto específico que interessa ao departamento envolvido, como vendas.

Segundo Kimball (1998, *apud* COLAÇO JUNIOR, 2004, p.18) “Um *Data Mart*, também é conhecido como *Warehouse* Departamental, é uma abordagem descentralizada do conceito de *Data Warehouse*”.

## 2.3 B.I. (Business Intelligence)

*Business Intelligence (BI)* é o objetivo da existência de um *data warehouse*, pois ferramentas de BI se utilizam dos dados contidos em um *data warehouse* para transformá-los em informações que fundamentarão as tomadas de decisão. De acordo com Turban



tempo de resposta das consultas será significativamente menor. A dimensão tempo foi criada com algumas descrições, o que permitiu a realização de consultas sem a utilização de funções do banco de dados, e esse fator tem influência positiva na performance das consultas.

O processo de ETL para carga dos dados no DW foi realizado através da geração de *scripts* SQL simples na base de dados de origem que foram executados na base DW para o carregamento dos dados, pois, apenas uma fonte de dados foi utilizada. Não foi necessária a utilização de ferramentas específicas de ETL devido a característica simples dos dados envolvidos nesse artigo.

Uma vez carregados os dados, foi possível a comparação das consultas que foram executadas em ambiente operacional com as consultas disponibilizadas no DW, sempre com os mesmos parâmetros, garantindo que os dados selecionados fossem exatamente os mesmos. Foram apontados os tempos de execução das consultas e também a sua complexidade baseada no número de ligações entre tabelas necessárias para a sua realização. Essas comparações permitiram demonstrar que o modelo dimensional apresenta vantagem sobre o modelo relacional em consultas de dados destinadas à tomada de decisão. Deve-se salientar que tanto as consultas realizadas ao modelo relacional quanto operacional foram executadas no mesmo servidor de testes, utilizando sempre a ferramenta de linha de comando, em tempos diferentes, procurando garantir as mesmas condições de uso, sem que nenhum outro processo estivesse concorrendo com as mesmas, e dessa forma, garantir qualidade nos resultados que serviram de base para as comparações.

Para o modelo dimensional proposto, foram definidas três tabelas de dimensão: *dim\_tempo*, *dim\_filial* e *dim\_cliente* e uma tabela de fato: *fato\_encomendas*. Após o processo de carga, essas tabelas apresentaram, respectivamente, 2.086, 322, 449.871 e 3.289.128 registros.

Foram realizadas ao todo nove consultas envolvendo níveis de complexidade diferentes, e que por isso, apresentaram variações de performance entre o modelo relacional e o modelo dimensional, mas sempre com números favoráveis ao modelo dimensional. Também foi considerado como fator de complexidade o número de ligações e também o número de funções de banco utilizadas.

Os resultados das consultas realizadas, mostrando a vantagem do modelo dimensional podem ser observados no quadro 2:

Quadro 2 – Comparação de resultados apresentados pelos modelos dimensional e relacional.

Objetivo da Consulta	Modelo Relacional			Modelo Dimensional		
	Tempo de execução(s)	Junções	Funções	Tempo de execução(s)	Junções	Funções
Mostrar o valor total de cargas do ano de 2014	13,41	0	2	1,23	1	1
Mostrar o valor total de cargas de 2013, totalizando mês a mês	15,81	0	2	1,55	1	1
Mostrar o valor total de cargas dos anos de 2010-2015, totalizando ano a ano	19,66	0	3	7,84	1	1
Mostrar o total financeiro bimestral no ano de 2013	25,76	0	3	2,97	1	1
Mostrar o total financeiro de 2013, totalizando por estado de origem	17,04	2	2	3,28	2	1
Mostrar o total financeiro de 2014, totalizando por estado de destino	22,71	2	2	7,77	2	1
Mostrar as 20 cidades de origem com maior valor de carga em 2010	17,04	2	3	3,60	2	1
Mostrar as 20 cidades de destino com maior valor de carga em 2010	22,38	2	3	8,83	2	1
Mostrar o valor total transportado no ano de 2012, totalizando por filial	18,29	1	3	8,96	2	1

Fonte: do autor

## 5. CONSIDERAÇÕES FINAIS

O modelo dimensional oferece vantagens em relação ao modelo relacional no que tange a performance e também a complexidade das consultas.



Em todas as consultas realizadas houve uma vantagem significativa de performance favorável ao modelo dimensional, e também uma simplicidade maior nas consultas, o que torna evidente a vantagem geral no caso da utilização dos dados para as tomadas de decisão.

O presente artigo considerou um modelo de Data Warehouse relativamente simples, embora com uma quantidade significativa de dados, devido ao tipo de granularidade utilizado, em um ambiente homogêneo no que diz respeito à origem dos dados e o foco foi justamente comprovar a vantagem do modelo dimensional em relação as consultas de dados. Em trabalhos futuros, o aprofundamento em modelos mais complexos e em ambientes heterogêneos podem revelar resultados diferentes e mais significativos em relação às vantagens do modelo dimensional, além de exigir processos mais elaborados no que diz respeito ao estudo dos dados de origem para a criação do modelo dimensional adequado e também o processo de ETL necessário para o mesmo.

## REFERENCIAS

ANTONIO, JOÃO. **Informática para concursos. 3.ed.** São Paulo:Elsevier, 2006.

BALTZAN, P.; PHILLIPS A. **Sistemas de Informação: A importância e as responsabilidades do pessoal de TI nas tomadas de decisões.** Porto Alegre: AMGH, 2012.

BARBIERI, C. BI – **Business Intelligence – Modelagem & Tecnologia.** Rio de Janeiro: Axcel Books, 2006.

BONEL, C. **Afinal, o que é Business Intelligence ?** 1.ed. Rio de Janeiro: Perse, 2015

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

DATE, C. J. **Introdução a Sistemas de Banco de Dados.** 8.ed. Rio de Janeiro: Elsevier, 2004.

FERREIRA, R.G.C. **Data Warehouse na prática: Fundamentos e Implantação,** Porto Alegre: Universidade Federal do Rio Grande do Sul, 2002.

HEUSER, C. A. **Projeto de Banco de Dados,** 4.ed. Porto Alegre: Sagra Luzzatto, 1998.

HOKAMA, D. D. B; CAMARGO, D; FUJITA, F; FOGLIENE, J. L.V. **A modelagem de dados no ambiente data warehouse.** São Paulo: Universidade Presbiteriana Mackenzie, 2004.

KAUARK, F. S.; MANHÃES, F. C.; MEDEIROS, C. H. **Metodologia da Pesquisa: Um Guia Prático,** Itabuna: Via Litterarum, 2010.

KIMBALL, R. **Data WareHouse Toolkit,** São Paulo: Makron Books, 1998.

KIMBALL, R; ROSS. M. **Data Webhouse: Construindo o Data Warehouse para a Web.** 1.ed. Rio de Janeiro: Campus, 2000

KIMBALL, R; ROSS. M. **The Data WareHouse Toolkit: The Complete guide to dimensional modeling.** 2.ed. New York: Wiley Computer Publishing, 2002.

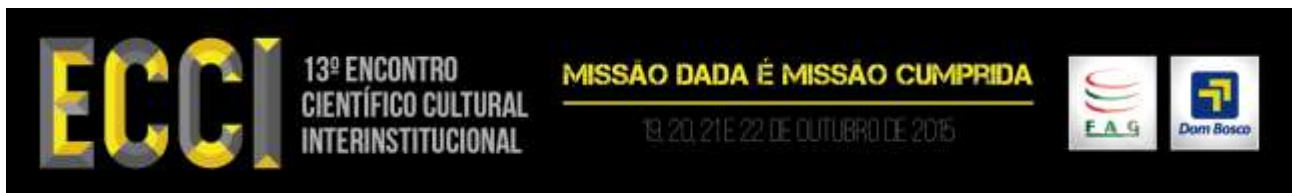
MACHADO, F. N. R. **Tecnologia e Projeto de Data Warehouse: Uma visão multidimensional.** Tatuapé: Érica, 2006.

MARCONI, M. A.; LAKATOS, E. M. **Técnicas de pesquisa: planejamento e execução de pesquisa, amostragens e técnicas de pesquisa, elaboração, análise e interpretação de dados.** 6. ed. São Paulo: Atlas, 2006.

MILANI, A. **Mysql guia do programador.** São Paulo: Novatec, 2006.

PRESTES, M. L. M. **A pesquisa e a construção do conhecimento científico: do planejamento aos textos, da escola à academia,** 3.ed. Catanduva: Rêspel, 2005.

PRIMAK, F. V. **Decisões com B.I. business intelligence.** São Paulo: Ciência Moderna, 2008.



RAMPAZZO, L. **Metodologia científica** para alunos dos cursos de graduação e pós-graduação. 3.ed. São Paulo: Loyola, 2005.

RASLAN, D. A.; CALAZANS, A. T. S. **Data Warehouse: conceitos e aplicações**, v.4, n.1, Brasília: Universitas Gestão e TI, 2014.

SASSI, R. J. **Análise das modalidades de um data warehouse (dw)**. v. 4. São Paulo: Sinergia, 2003.

SILBERSCHATZ, A.; KORTH, H. F; SUDARSHAN, S. **Sistema de Banco de Dados**. 3.ed. São Paulo: Makron Books, 2004.

SINGH, H. S. **Data Warehouse: conceitos, tecnologias, implementação e gerenciamento**. São Paulo: Makron Books, 2001.

TURBAN E.; SHARDA. R.; ARONSON J. E.; KING D. **Business intelligence: um enfoque gerencial para a inteligência do negócio**, 1.ed. Porto Alegre: Artmed, 2008.

VELLOSO F. C. **Informática: Conceitos básicos**: 8.ed. São Paulo: Elsevier, 2011.