

Efraim Turban

Ramesh Sharda

Jay E. Aronson

David King

BUSINESS INTELLIGENCE

UM ENFOQUE GERENCIAL PARA A
INTELIGÊNCIA DO NEGÓCIO





B979 Business intelligence [recurso eletrônico] : um enfoque gerencial para a inteligência do negócio / Efraim Turban ...[et al.] ; tradução Fabiano Bruno Gonçalves. – Dados eletrônicos. – Porto Alegre : Bookman, 2009.

Editado também como livro impresso em 2009.
ISBN 978-85-7780-425-2

Administração de empresas. I. Turban, Efraim.

CDU 658

Efraim Turban

Universidade do Havái

Ramesh Sharda

Universidade Estadual de Oklahoma

Jay E. Aronson

Universidade da Geórgia

David King

JDA Software Group, Inc.

BUSINESS INTELLIGENCE

UM ENFOQUE GERENCIAL PARA A
INTELIGÊNCIA DO NEGÓCIO

Tradução:

Fabiano Bruno Gonçalves

Consultoria, supervisão e revisão técnica desta edição:

Ângela Freitag Brodbeck, PhD

Escola de Administração

Grupo de Pesquisa em Sistemas de Informação e de Apoio à Decisão

Universidade Federal do Rio Grande do Sul

Versão impressa

desta obra: 2009



2009

Obra originalmente publicada sob o título
Business Intelligence: a managerial approach, 1st ed.

ISBN 978-0-13-234761-7

© 2008 by Pearson Education, Inc.

Tradução autorizada a partir do original em língua inglesa publicado por Pearson Education, Inc., sob o selo de Prentice Hall.

Capa: *Gustavo Demarchi*

Leitura final: *Vinicius Selbach e Monica Stefani*

Supervisão editorial: *Arysinha Jacques Affonso*

Editoração eletrônica: *Techbooks*

Reservados todos os direitos de publicação, em língua portuguesa, à
ARTMED® EDITORA S.A.
(BOOKMAN® COMPANHIA EDITORA é uma divisão da ARTMED® EDITORA S.A.)
Av. Jerônimo de Ornelas, 670 - Santana
90040-340 Porto Alegre RS
Fone (51) 3027-7000 Fax (51) 3027-7070

É proibida a duplicação ou reprodução deste volume, no todo ou em parte,
sob quaisquer formas ou por quaisquer meios (eletrônico, mecânico, gravação,
fotocópia, distribuição na Web e outros), sem permissão expressa da Editora.

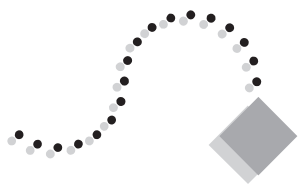
SÃO PAULO
Av. Angélica, 1091 - Higienópolis
01227-100 São Paulo SP
Fone (11) 3665-1100 Fax (11) 3667-1333

SAC 0800 703-3444

IMPRESSO NO BRASIL
PRINTED IN BRAZIL

Dedicado a todos aqueles que se interessam em aprender sobre BI – business intelligence.

—Os autores



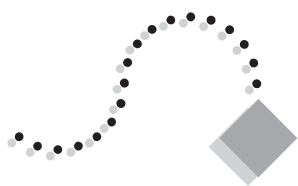
Os Autores

Efraim Turban (M.B.A., Ph.D, Universidade da Califórnia, Berkeley) é professor convidado no Pacific Institute for Information System Management, na Universidade do Havaí. Antes disso, trabalhou em diversas universidades, incluindo City University of Hong Kong; Lehigh University; Florida International University; California State University, Long Beach; Eastern Illinois University, Long Beach e University of Southern California. Turban é autor de mais de 100 artigos, publicados em periódicos como o *Management Science*, o *MIS Quarterly* e o *Decision Support Systems*. É também autor de 21 livros, entre os quais se incluem *Electronic Commerce: A Managerial Perspective*, *Decision Support Systems* e *Tecnologia da Informação para Gestão*. Ele também é consultor de grandes corporações mundiais. Suas atuais áreas de interesse são sistemas de suporte a decisões baseados na Web, o uso de agentes inteligentes em sistemas de comércio eletrônico e questões de colaboração no comércio eletrônico global.

Ramesh Sharda (M.B.A., Ph.D., University of Wisconsin-Madison) é diretor do Institute of Research in Information Systems (IRIS), Presidente de Gestão de Tecnologia da ConocoPhillips e Regents Professor de Ciência da Administração e Sistemas de Informação na Spears School of Business Administration na Oklahoma State University (OSU). Ele iniciou e atuou como diretor do M.S., um Programa de Gerenciamento de Telecomunicações na OSU. Mais de 100 artigos que descrevem sua pesquisa já foram publicados nos principais periódicos, incluindo *Management Science*, *Information Systems Research*, *Decision Support Systems* e *Journal of Information Systems*. Sharda colabora em diversos conselhos editoriais, entre os quais do *INFORMS Journal of Computing*, do *Decision Support System* e do *Journal of Management Frontiers*. Seus interesses atuais de pesquisa são em sistemas de suporte a decisões, em aplicações colaborativas e em tecnologias para gerenciar a sobrecarga de informações. Ele é co-fundador da iTradeFair.com, uma empresa que produz feiras virtuais de negócios.

Jay E. Aronson (M.A., M.S., Ph.D., Carnegie Mellon University) é professor de sistemas de informações gerenciais no Terry College of Business, na University of Georgia. Fez parte do corpo docente da Southern Methodist University. Aronson é autor de cerca de 50 artigos publicados em periódicos como o *Management Science*, o *Information Systems Research* e o *MIS Quarterly*. É autor de três livros e contribuiu com várias enciclopédias profissionais. Também é consultor de grandes corporações e organizações internacionais. As atuais áreas de pesquisa do Dr. Aronson incluem a gestão do conhecimento, a computação colaborativa e a computação paralela.

Dave King (Ph.D) tem mais de 25 anos de experiência em liderança do desenvolvimento do suporte a decisões, gerenciamento de desempenho e *software* empresarial. É vice-presidente sênior da New Product Development, na JDA Software, Inc., em Scottsdale, no Arizona. Ele se uniu à JDA em 2004, após vários anos como vice-presidente sênior de Desenvolvimento de Produtos e CTO da Comshare Inc. King é autor de vários artigos e livros e é co-autor do *Electronic Commerce: A Managerial Perspective* (Prentice Hall). Ele também participa de diversos conselhos consultivos e de universidades, como MIS Advisory, na University of Georgia, e o Technopolis Advisory Board, na Arizona State University.



Prefácio

À medida que começamos o século 21, vemos grandes mudanças na forma como a gerência das empresas usa o computador para tomar decisões. Cada vez mais tomadores de decisão têm conhecimento sobre computadores e a Web, fazendo uso de seus recursos, no seu trabalho. O campo do *business intelligence* (BI) está evoluindo a partir de sua origem como, principalmente, ferramenta para executivos e se torna rapidamente uma *commodity* compartilhada por gerentes, analistas e altos executivos. Hoje em dia, as empresas podem usar facilmente intranets e a Internet para fornecer aplicativos de análise de desempenho, de grande valor, para os tomadores de decisão ao redor do mundo. As corporações desenvolvem regularmente sistemas distribuídos, intranets e extranets, que possibilitam o fácil acesso a dados armazenados em locais múltiplos que permitem a colaboração e comunicação em todo o globo. Vários aplicativos de sistemas de informação são integrados entre si e/ou com outros sistemas baseados na Web. Algumas integrações até mesmo transcendem os limites empresariais. Os administradores podem tomar melhores decisões, pois dispõem de informações mais precisas e oportunas na ponta dos dedos.

Business intelligence (BI) é um termo “guarda-chuva” que engloba ferramentas, arquitetura, bases de dados, *data warehouse*, gerenciamento de desempenho, metodologias e assim sucessivamente, tudo integrado em uma suíte de *software*. O objetivo do *software* é de possibilitar que os gerentes de negócios e analistas em uma empresa acessem qualquer dado da empresa de maneira fácil e rápida, possivelmente em tempo real, bem como conduzir manipulações e análises apropriadas. Analisando dados históricos e atuais, situações, métricas e desempenhos, os tomadores de decisões obtêm valiosos *insights* que os ajudam a tomar melhores decisões. O BI possui várias capacidades, o que inclui relatórios e perguntas, análise complicada, *data mining*, previsões e muito mais. Essas capacidades vieram de ferramentas e tecnologias nas quais o BI se baseia, e especialmente de sistemas de informação executiva (EIS), sistemas de apoio à decisão (DSS), perguntas, visualização, fluxo de trabalho, ciência de pesquisa/gerenciamento de operações e inteligência artificial aplicada.

O BI utiliza os poderosos computadores de hoje, bem como as redes, a Internet e outras plataformas para elevar essas (e outras) tecnologias ao nível mais alto possível. Essas tecnologias são integradas com outras ferramentas (p. ex., *software* ERP e CRM) de tal forma que se tornam da maior utilidade para todos os usuários. Por tal razão, quase todas médias e grandes empresas hoje em dia estão utilizando BI, ou alguns módulos do BI, para melhorar seu desempenho, se superar, e, por vezes, até mesmo sobreviver.

Hoje em dia, os usuários de BI utilizam a Web para a sua análise, usam interfaces gráficas que permitem que as tomadas de decisões sejam flexíveis, eficientes e fáceis de visualizar, também que processem dados e modelos usando navegadores da Web familiares. Os gerentes podem também se comunicar com computadores e a Web usando uma variedade de dispositivos portáteis sem fio, o que inclui celulares e PDAs. Esses dispositivos permitem que os gerentes acessem informações importantes e ferramentas úteis para comunicação e colaboração. O *data warehouse* e suas ferramentas analíticas [p. ex., processamento analítico *online* (OLAP), *data mining*] aumentam muito o acesso e a análise de informações entre as fronteiras das empresas.

O objetivo deste livro é apresentar ao leitor essas várias tecnologias. Ele apresenta os passos fundamentais das técnicas de BI e a maneira pela qual os sistemas de BI são construídos e utilizados. BI para o apoio de decisões empresariais é o tema deste livro que aumenta o entendimento do leitor em relação ao mundo da Web, fornecendo exemplos, produtos, serviços, além de exercícios, e discutindo problemas relacionados à Internet. O livro tem o apoio de suplementos apresentados como arquivos *online* (www.bookman.com.br).

Esta edição foi planejada como um livro-texto para um curso de Inteligência de Negócios, bem como um texto de apoio para cursos como Introdução a SIG ou Estratégia de Negócios. Pode também suplementar um curso de Gerenciamento de Tecnologias para um MBA – qualquer curso que lide com desempenho organizacional e seu gerenciamento será beneficiado pelo seu uso.



RECURSOS

Muitos recursos são exclusivos deste livro.*

Orientação gerencial

Pode-se focar a inteligência de negócios a partir de dois principais pontos de vista: tecnológico e gerencial. Este texto toma o segundo enfoque. A maioria das apresentações são sobre aplicações e implementação do BI. Entretanto, reconhecemos a importância da tecnologia e fornecemos diversos arquivos e Capítulo 6 de forma *online*, no site da Bookman Editora (www.bookman.com.br).

Orientação no mundo real

Exemplos extensos e vívidos de grandes empresas, pequenas empresas, governos e agências sem fins lucrativos de todo o mundo tornam os conceitos mais reais. Esses exemplos mostram aos alunos as capacidades do BI, seus custos e justificativas, e as maneiras inovadoras com que as empresas estão usando o BI nas suas operações.

A conexão da Teradata University Network (TUN)

O TUN é um portal de aprendizado gratuito patrocinado pela Teradata, uma divisão da NCR, cujo objetivo é ajudar os membros dos corpos docentes a aprenderem, ensinarem, comunicarem e colaborarem com outros no campo do BI. Várias centenas de universidades e corpos docentes participam do TUN e o usam. A Teradata também apóia um portal para estudantes (teradatastudentnetwork.com), que contém uma quantidade considerável de recursos de aprendizado, como casos, seminários da Web, tutoriais, exercícios, *links* para fontes e muito mais. Nosso texto é interconectado com a TUN, principalmente por meio das várias atribuições em todos os capítulos oferecidos aos alunos pelo portal.

Tópicos mais atuais

O livro apresenta os tópicos mais atuais em relação ao BI, como se vê pelas muitas citações de 2005 e 2006.

Sistemas integrados

Em comparação a outros livros que destacam sistemas isolados de BI com base na Internet, enfatizamos os sistemas que apóiam a empresa e seus muitos usuários.

* N. de R.: Alguns dos recursos aqui mencionados estão disponíveis apenas em língua inglesa. Por favor, consulte o site da Bookman Editora (www.bookman.com.br) para verificar os recursos disponíveis em português.

Perspectiva global

A importância da concorrência global, parcerias e comércio está crescendo rapidamente. A CE facilita a importação e exportação, o gerenciamento de empresas multinacionais e o comércio eletrônico ao redor do mundo. Exemplos internacionais são fornecidos ao longo do livro.

Suporte online

Arquivos estão disponíveis *online* para suplementar o texto. Esse material inclui arquivos sobre tópicos genéricos, tais como *data mining* e *intranets*, casos, texto orientado tecnicamente e muito mais.

Tutoriais

Dois principais tutoriais são fornecidos *online*.

Amigável ao usuário

Ao mesmo tempo que cobre todos os principais tópicos do BI, este livro é claro, simples e bem-organizado. Ele fornece todas as definições básicas dos termos, bem como apoio lógico conceitual. Além disso, o livro é fácil de entender e está cheio de exemplos interessantes do mundo real que mantêm o interesse do leitor em alto nível. Questões relevantes para revisão são fornecidas no final de cada seção, de modo que o leitor pode parar para revisar e digerir o novo material.

AUXÍLIOS NO APRENDIZADO

O texto fornece vários auxílios no aprendizado do aluno:

- **Visão geral.** Uma lista dos principais títulos do capítulo no seu início fornece uma visão geral rápida dos principais tópicos abordados.
- **Objetivos de aprendizado.** Os objetivos de aprendizado no início de cada capítulo ajudam o aluno a concentrar seus esforços e os alertam para os conceitos importantes a serem discutidos.
- **Vinhetas de abertura.** Cada capítulo inicia com um exemplo do mundo real que ilustra a importância do BI para a empresa moderna. Esses casos foram escolhidos com cuidado para chamar a atenção para os principais tópicos abordados nos capítulos. Perguntas seguem cada caso, de modo a auxiliar a direcionar a atenção do estudante para as implicações do material do caso. Após cada vinheta, uma seção curta intitulada “O que podemos aprender...” vincula as questões importantes da vinheta com o assunto do capítulo.
- **Casos de aplicação do BI.** Casos apresentados no capítulo destacam problemas do mundo real encontrados pela empresas enquanto elas desenvolvem e implementam o BI.
- **Insights de tecnologia.** Às vezes, os tópicos exigem mais elaboração ou demonstração. Quadros com *insights* e adições fornecem um repositório atraente para esse conteúdo.
- **Exibições.** Muitas exibições atraentes (figuras e tabelas) aumentam e suplementam a discussão do texto. Muitas delas estão disponíveis *online*.
- **Questões de revisão.** Cada seção é finalizada com uma série de questões de revisão sobre a mesma. Com essas questões, tem-se a intenção de ajudar os alunos a resumir os conceitos apresentados e digerir as partes essenciais de cada seção antes de avançar para outro tópico.

- **Termos-chave.** Cada termo-chave é definido no texto e registrado numa lista de termos-chave no final de cada capítulo. Além disso, um glossário alfabético de termos-chave aparece no final do livro.
- **Questões gerenciais.** No final de cada capítulo, exploramos algumas das preocupações especiais que os gerentes enfrentam enquanto se adaptam a fazer negócios no espaço cibernético. Essas questões são enquadradas como perguntas, de forma a maximizar o compromisso ativo dos leitores com elas.
- **Destaques dos capítulos.** Estes destaques resumem os principais tópicos apresentados em cada capítulo.
- **Exercícios no final do capítulo.** Diferentes tipos de questões medem a compreensão do aluno e sua capacidade de aplicar o conhecimento. Questões para Discussão têm a intenção de promover o debate em aula, além de desenvolver habilidades de pensamento crítico. Exercícios na Internet são atribuições desafiadoras que exigem que os alunos efetuem buscas e apliquem o que aprenderam. Mais de 50 exercícios práticos remetem os alunos para *sites* interessantes na Web, de modo a conduzir a pesquisa, investigar um aplicativo, efetuar o *download* de demos ou aprender sobre a tecnologia mais avançada. Os exercícios de Atribuição de Grupo e de Assumir Papéis são projetos desafiadores planejados para fomentar o trabalho em grupo.
- **Casos reais.** Cada capítulo termina com um caso real, apresentado de uma maneira mais aprofundada do que os Casos de Aplicações do BI nos capítulos. Há questões após cada caso.
- **Exercícios da Teradata University e outros exercícios práticos.** Muitas tarefas exigem o uso do portal do aluno da TUN.



MATERIAIS SUPLEMENTARES EM WWW.BOOKMAN.COM.BR

Os seguintes materiais estão disponíveis a todos os leitores cadastrados no *site* da Bookman Editora:

- um capítulo *on line* (Capítulo 6, sobre Computação Neural e *Data Mining*)
- Dois tutoriais de tecnologia
- Arquivos *on line* de todos os capítulos, com apresentações técnicas, casos e tabelas, organizados de acordo com a sequência em que aparecem no capítulo

Já os professores que adotarem esta obra poderão valer-se dos seguintes recursos, **todos em língua inglesa**:

- Manual do professor, com todas as respostas a questões de revisão e discussão, exercícios e questões do caso de aplicação
- Arquivo de item de teste: um amplo conjunto de questões de múltipla escolha, verdadeiro ou falso, e de ensaio para cada capítulo
- Apresentações PowerPoint, voltadas para objetivos de aprendizado do texto
- Biblioteca de imagens publicadas no livro

Os tutoriais disponíveis para *download* no *site* da Bookman foram concebidos para utilização com o *software* STATISTICA. Salientamos que o *software*, e demais arquivos de dados de amostras relacionados a ele, não são oferecidos com este livro e os interessados em adquiri-los deverão entrar em contato com os representantes no Brasil, a StatSoft, localizada em São Paulo, pelo endereço www.statsoft.com.br para mais detalhes.



AGRADECIMENTOS

Muitas pessoas nos ajudaram a produzir este texto. O retorno de corpos docentes foi solicitado por meio de revisões e um grupo de foco. Somos gratos aos seguintes corpos docentes por suas contribuições.

Contribuições por conteúdo

As seguintes pessoas contribuíram com material para a edição:

- Michel Goul, da Arizona State University, forneceu material para o Capítulo 1.
- Leila Halawi, da Nova Southeastern University, forneceu material para o Capítulo 2.
- Ting-Peng Liang, da National Sun Yet Sen University, em Taiwan, forneceu material para o Capítulo 4 e o Capítulo 6 *online*.
- Subramanian Rama Iyer contribuiu com os estudos de caso no Capítulo 4 e no Capítulo 6 *online*.

Revisores

Desejamos agradecer ao corpo docente que participou de revisões do texto e do nosso outro título DSS:

Ranjit Bose, University of New Mexico
 Lee Roy Bronner, Morgan State University
 Wingyan Chung, University of Texas
 Jerry Fjermestad, New Jersey Institute of Technology
 Yair Levy, Nova Southeastern University
 Joshua Pauli, Dakota State University
 Peter Sientins, Bridgewater State College
 Charles Stout, High Point University
 Meenu Singh, Murray State University
 Howard (Charles) Walton, Gettysburg University

Reconhecimento e agradecimentos especiais

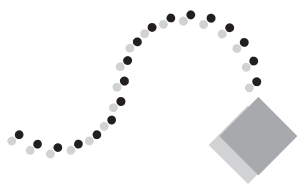
Agradecimentos especiais à Teradata University Network e especialmente ao seu diretor sênior, Hugh Watson; a Mary Gros, diretora de relações acadêmicas; à Teradata, uma divisão da NCR e a uma conexão entre a Teradata e a comunidade acadêmica, por encorajarem vincular este livro com a TUN e por fornecerem materiais úteis para o livro.

O trabalho de tantos escritores, contribuidores e participantes foi gerenciado com tanta habilidade por Judy Lang, que colaborou com todos nós, preparou a edição e nos guiou durante todo o projeto. Judy também preparou os slides de PowerPoint.

Por fim, à equipe da Prentice Hall, que deve ser louvada: Bob Horan, Editor Executivo, que guiou este projeto; Ana Jankowski, Editora Assistente; Kelly Loftus, Assistente Editorial, Laura Cirigliano, Assistente de *Marketing*, e Ashley Lulling, nossa Gerente de Projetos de Mídia. Também gostaríamos de agradecer à equipe de produção que inclui Carol Samet e Heidi Allgair, da GGS Book Services, que transformou o manuscrito em um livro.

Gostaríamos de agradecer a todas essas pessoas e empresas. Sem sua ajuda, a criação deste livro não teria sido possível.

E.T.
 R.S.
 J.E.A.
 D.K.



Sumário Resumido

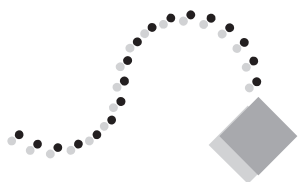
CAPÍTULO 1	Introdução ao <i>Business Intelligence</i>	21
CAPÍTULO 2	<i>Data warehousing</i>	54
CAPÍTULO 3	Análise de negócios e visualização de dados	101
CAPÍTULO 4	<i>Data, text</i> e <i>web mining</i>	150
CAPÍTULO 5	<i>Business Performance Management</i> (BPM)	191

MATERIAL ONLINE

CAPÍTULO 6	Redes neurais para <i>data mining</i>	239
-------------------	---	-----

TUTORIAL ONLINE

Tutorial 1	Projeto de <i>text mining</i>	T1-1
Tutorial 2	Projeto do <i>software</i> Statistica	T2-1




Sumário

CAPÍTULO 1	Introdução ao <i>Business Intelligence</i>	21
1.1	Vinheta de abertura: a Toyota usa o <i>Business Intelligence</i> para se superar	22
1.2	Alterando os ambientes de negócios e o suporte computadorizado à decisão	24
1.3	Uma estrutura para o <i>Business Intelligence</i> (BI)	26
	<i>Determinação inteligente de preços usando o suporte automatizado à decisão</i>	30
	<i>Análise preditiva ajuda o Texas a arrecadar impostos</i>	33
1.4	Criação e uso da inteligência e governança do BI	34
1.5	As principais teorias e características do <i>Business Intelligence</i>	36
1.6	Rumo à inteligência e à vantagem competitivas	39
1.7	A implementação bem-sucedida do <i>Business Intelligence</i>	42
	<i>O Business Intelligence da France Telecom</i>	44
1.8	Conclusão: o <i>Business Intelligence</i> hoje e amanhã	47
1.9	Recursos, links e conexão com a Teradata University Network	48
	<i>BNSF usa BI para melhorar serviço, agilidade, eficiência e lucratividade</i>	51
	Referências	52
CAPÍTULO 2	<i>Data warehousing</i>	54
2.1	Vinheta de abertura: a Continental Airlines voa alto com seu <i>data warehouse</i> em tempo real	54
2.2	Definições e conceitos de <i>data warehousing</i>	57
2.3	Visão geral do processo de <i>data warehousing</i>	60
	<i>O data warehousing apóia a estratégia corporativa da First American Corporation</i>	60
2.4	Arquiteturas de <i>data warehousing</i>	62
2.5	Integração de dados e processos de Extração, Transformação e Carga (ETL)	70
	<i>O data warehouse integrado e premiado do Bank of America</i>	71
2.6	Desenvolvimento de <i>data warehouses</i>	74
	<i>As coisas melhoram com o data warehouse da Coca-Cola</i>	75
	<i>HP consolida centenas de data marts em um único EDW</i>	79
2.7	<i>Data warehousing</i> em tempo real	86
	<i>O Egg plc arrasa a concorrência quase em tempo real</i>	87

2.8	Questões de administração e segurança de <i>data warehouses</i>	91
2.9	Recursos, <i>links</i> e conexão com a Teradata University Network	92
	<i>Data warehousing em tempo real na Overstock.com.</i>	97
	Referências	98
CAPÍTULO 3	Análise de negócios e visualização de dados	101
3.1	Vinheta de abertura: a Lexmark International aperfeiçoa as operações com o <i>Business Intelligence</i>	102
3.2	O campo da análise de negócios (BA): visão geral	104
	<i>Ben & Jerry's se supera com a BA</i>	105
3.3	Processamento analítico <i>online</i> (OLAP)	109
	<i>TCF Financial Corp.: conduzindo OLAP, relatórios e data mining</i>	113
3.4	Relatórios e consultas	114
3.5	Multidimensionalidade	117
3.6	Análise de negócios avançada.	120
	<i>A análise preditiva pode ajudá-lo a evitar engarrafamentos.</i>	121
3.7	Visualização de dados	124
	<i>A visualização de dados financeiros na Merrill Lynch</i>	127
3.8	Sistemas de informação geográfica (GIS).	127
	<i>GIS e GPS rastreiam onde você está e ajudam a fazer as coisas</i>	130
3.9	<i>Business Intelligence</i> em tempo real, suporte automatizado à decisão (ADS) e inteligência competitiva	131
3.10	Análise de negócios e a Web: inteligência da Web e análise da Web	137
	<i>Análise da Web aperfeiçoa o desempenho do comércio eletrônico</i>	139
3.11	Uso, benefícios e sucesso da análise de negócios	140
	<i>Varejistas tornam estável o progresso do BI.</i>	142
	<i>Governos estaduais compartilham informações geoespaciais.</i>	147
	Referências	148
CAPÍTULO 4	<i>Data, text e web mining</i>	150
4.1	Vinheta de abertura: Highmark Inc. emprega <i>data mining</i> para gerenciar as despesas com seguros	150
4.2	Conceitos e aplicações de <i>data mining</i>	152
	<i>Dados ajudam a prever as necessidades do cliente</i>	154
	<i>Acidentes com veículos automotores e as distrações do motorista.</i>	157
	<i>Data mining para identificar o comportamento do cliente.</i>	158
	<i>A personalização da medicina.</i>	160
	<i>Uma mina no financiamento de terroristas</i>	160
4.3	Técnicas e ferramentas de <i>data mining</i>	161
4.4	Processos do projeto de <i>data mining</i>	173
	<i>Desdobramentos e avanços do data mining do DHS no cumprimento da lei</i>	176

4.5	<i>Text mining</i>	177
	<i>Voando pelo texto</i>	178
4.6	<i>Web mining</i>	181
	<i>Apanhado pela rede</i>	182
	<i>Hewlett-Packard e text mining</i>	188
	<i>Referências</i>	190
CAPÍTULO 5	<i>Business Performance Management (BPM)</i>	191
5.1	Vinheta de abertura: Cisco e o fechamento virtual	192
5.2	Visão geral do <i>Business Performance Management (BPM)</i>	194
5.3	Estratégia: aonde queremos ir?	196
5.4	Plano: como chegaremos lá?	198
5.5	Monitoração: como estamos fazendo?	200
	<i>Planejamento dirigido por descoberta: o caso da Euro Disney</i>	202
5.6	Ação e ajuste: o que precisamos fazer de forma diferente?	203
5.7	Medida de desempenho	206
	<i>International Truck and Engine Corporation</i>	208
5.8	Metodologias de BPM	210
5.9	Arquitetura e aplicações de BPM	217
5.10	<i>Dashboards</i> de desempenho	225
	<i>Dashboards para médicos</i>	227
5.11	Monitoramento de atividades de negócios (BAM)	229
	<i>A cidade de Albuquerque em tempo real</i>	230
	<i>Sistemas de informação vigilantes na Western Digital</i>	236
	<i>Referências</i>	238
MATERIAL ONLINE		
CAPÍTULO 6	<i>Redes neurais para data mining</i>	239
Capítulo 1	<i>Arquivos Online</i>	W1-1
W1.1	O processo geral de criação e uso de inteligência conforme se reflete na Continental Airlines (nota: o caso é apresentado no Capítulo 2 e está disponível no <i>Website</i> da TUN.)	W1-1
W1.2	Controle do BI	W1-3
W1.3	A comunidade de usuários de BI	W1-4
W1.4	Um plano de ação para a organização de SI que faz projetos de DW/BI	W1-6
W1.5	Teradata University Network (TUN): um recurso importante para professores de <i>data warehousing</i> , DSS/BI e bancos de dados	W1-7
Capítulo 3	<i>Arquivos Online</i>	W3-1
W3.1	Principais capacidades do EIS	W3-1
W3.2	SAP Analytics	W3-2
W3.3	Tendências em produtos de visualização para o suporte à decisão	W3-3
W3.4	Visualização da realidade virtual	W3-4

W3.5	Inteligência competitiva na Internet	W3-5
W3.6	Cabela's	W3-6
Capítulo 4	<i>Arquivos Online</i>	W4-1
W4.1	<i>Data mining</i>	W4-1
Capítulo 5	<i>Arquivos Online</i>	W5-1
W5.1	Portfólio de opções.	W5-1
W5.2	Previsões contínuas e dados em tempo real.	W5-2
W5.3	Medição de desempenho eficaz	W5-2
W5.4	Funções do Seis Sigma	W5-4
W5.5	Problemas com mostradores de <i>dashboard</i>	W5-5
 TUTORIAL ONLINE		
Tutorial 1	Projeto de <i>text mining</i>	T1-1
Tutorial 2	Projeto do <i>software</i> Statistica	T2-1
 GLOSSÁRIO 241		
ÍNDICE 245		

A decorative graphic on the left side of the page, consisting of a series of small squares arranged in a curved, descending line from the top left towards the bottom left. The squares are in two shades of gray.

BI: UM ENFOQUE GERENCIAL PARA A INTELIGÊNCIA DO NEGÓCIO

Neste livro, embarcamos na descoberta de um fenômeno estimulante nas empresas modernas: esforços emergentes para usar o *business intelligence* para melhorar a tomada de decisões, aperfeiçoar a posição estratégica e sustentar a vantagem competitiva.



Introdução ao *Business Intelligence*

Objetivos de aprendizado

- ◆ Entender o ambiente turbulento dos negócios e descrever como as empresas sobrevivem e, até mesmo, se superam nesse ambiente (resolvendo problemas e explorando oportunidades)
- ◆ Entender a necessidade do apoio de computadores nas tomadas de decisões gerenciais
- ◆ Descrever a metodologia e os conceitos do *business intelligence* (BI) e relacioná-los ao DSS
- ◆ Entender as questões principais ao implementar o *business intelligence*

O ambiente de negócios no qual as empresas operam atualmente está se tornando cada vez mais complexo e mutante. As empresas, privadas ou públicas, sentem crescentes pressões forçando-as a responder rapidamente a condições que estão sempre mudando, além de terem que ser inovadoras na maneira com que operam. Essas atividades exigem das empresas agilidade, tomadas de decisão rápidas e frequentes, sejam elas estratégicas, táticas e operacionais, algumas das quais são muito complexas. Tomar essas decisões pode exigir quantidades consideráveis de dados oportunos e relevantes, além de informações e conhecimento. O processamento dessas informações, na estrutura das decisões necessárias, deve ser feito de forma rápida, com frequência em tempo real, e comumente exige algum apoio computadorizado.

Este livro trata do uso do *business intelligence* como um apoio computadorizado para a tomada de decisões gerencial. Ele se concentra nos fundamentos teóricos e conceituais desse apoio, bem como nas ferramentas e técnicas comerciais que estão disponíveis. Este capítulo introdutório fornece mais detalhes sobre os tópicos acima, além de uma visão geral do livro.

As seções específicas são:

- 1.1 Vinheta de abertura: a Toyota usa o *business intelligence* para se superar
- 1.2 Alterando os ambientes de negócios e o suporte de decisões computadorizado
- 1.3 Uma estrutura para o *business intelligence* (BI)
- 1.4 Criação de inteligência e seu uso, e o controle da BI
- 1.5 As principais teorias e características do *business intelligence*
- 1.6 Rumo à inteligência e vantagem competitivas
- 1.7 A implementação bem-sucedida do *business intelligence*
- 1.8 Conclusão: o *business intelligence* hoje e amanhã
- 1.9 Recursos, *links* e a conexão da Teradata University Network

VINHETA DE ABERTURA: A TOYOTA USA O *BUSINESS INTELLIGENCE* PARA SE SUPERAR

O problema

A Toyota Motor Sales USA (**toyota.com**) é a distribuidora norte-americana de carros e caminhões construídos pela Toyota (uma subsidiária da Toyota). A empresa compra os carros na fábrica da Toyota no Japão e em outros locais, toma posse dos veículos e então os vende para os revendedores da Toyota nos Estados Unidos. Um veículo de médio porte custa US\$ 8 por dia enquanto está em trânsito. Como era de costume levar de 9 a 10 dias em trânsito, acabava por custar de US\$ 72 a 80 por carro. Para dois milhões de carros por ano, o custo para a empresa era de US\$ 144 a 160 milhões por ano. Isso era demais.

No final dos anos 90, a empresa enfrentou problemas cada vez maiores na sua cadeia de fornecimento e em suas operações, e os seus custos para armazenar carros se elevaram. Além disso, a incapacidade de fornecer carros para os revendedores a tempo resultou em clientes infelizes, que então compravam carros dos concorrentes, como a Honda. Isso se tornou extremamente importante em 2003 e 2004, quando carros híbridos foram apresentados e a concorrência com a Honda se intensificou.

No passado, a gerência usava computadores que geravam “toneladas” de relatórios e dados sem direção. A gerência não tinha a capacidade de usar esses dados e relatórios estrategicamente. Além disso, departamentos internos geralmente não tinham sucesso em compartilhar informações ou o faziam muito devagar. Relatórios acionáveis eram comumente produzidos tarde demais. Além disso, sistemas de relatórios sobrepostos forneciam dados que não eram sempre exatos. A gerência não era capaz de tomar decisões em tempo hábil, já que não era certo quais as porções dos dados que tinham precisão. A situação era especialmente penosa na divisão dos Serviços Logísticos da Toyota (TLS), que gerenciava o transporte dos veículos.

A gerência da TLS exige rastreamento preciso e gerenciamento de cadeia de fornecimento para assegurar que os carros certos vão para os revendedores certos a tempo. O agendamento manual e outros processos relacionados com negócios que foram conduzidos com informações incorretas causaram mais problemas. Por exemplo, se uma pessoa ocasionou um erro de entrada de dados quando um navio chegou ao estaleiro, o erro persistirá em toda a cadeia de fornecimento. (Por exemplo, alguns dados indicaram à gerência que os navios nunca chegaram a um porto semanas após os navios chegarem seguramente ao estaleiro.) As empresas de TI não foram capazes de responder à crescente necessidade do negócio. Por fim, um novo CIO foi contratado em 1997 para consertar os problemas.

A solução

Barbara Cooper, a nova CIO, iniciou tentando identificar os problemas exatos. Uma coisa se tornou clara: era necessário um *data warehouse*. Um *data warehouse* é um repositório central de dados históricos, organizado de forma que o acesso seja fácil (usando um navegador da Web) e a manipulação para o suporte a decisões é conveniente (veja a discussão mais adiante neste capítulo e no Capítulo 2). Também se tornou claro que ferramentas de *software* para efetuar o processamento, a exploração, e manipulação de dados eram necessárias. Foi configurado um sistema para fornecer dados precisos e em tempo real. Infelizmente, o sistema não funcionou apropriadamente. Em primeiro lugar, a entrada de dados históricos no sistema incluía anos de erros humanos que passaram despercebidos, o que incluía dados duplicados inconsistentes, bem como dados que faltavam. Isso produziu resultados e análises errôneas. Dessa forma, o novo sistema carecia de capacidades de fornecer o que os gerentes precisavam. Em 1999, tornou-se claro que a solução não funcionou. Era um conceito correto, mas com tecnologia errada dos fornecedores errados. Em 2000, a Toyota mudou para uma

tecnologia melhor. Usando o *data warehouse* da Oracle e a plataforma do *business intelligence* da Hyperion, foi criado um novo sistema. O sistema também incluía o recurso de *dashboard*, da Hyperion (Capítulos 2 e 5), que permite que os executivos vejam áreas que merecem atenção em suas unidades de negócio e investiguem mais para identificar os problemas com exatidão, bem como as suas causas. Usando cores diferentes (p. ex., vermelho para perigo), um gerente de negócios pode ver em tempo real, por exemplo, quando os tempos de entrega estão se tornando vagarosos e encontrar imediatamente as fontes do problema e até mesmo avaliar soluções potenciais do tipo “e se...”.

Os resultados

Dentro de poucos dias, o sistema começou a apresentar resultados fascinantes. Por exemplo, o sistema ajudou a descobrir que a Toyota era cobrada duas vezes por um envio especial por trem (um erro de US\$ 800.000). No geral, a Toyota USA conseguiu aumentar o volume de carros que negociava em 40% entre 2001 e 2005, enquanto aumentou o número de funcionários em apenas 3%. Além disso, o tempo de trânsito foi reduzido em mais de 5%. A fórmula do sucesso da TLS com seu BI rapidamente se espalha pela Toyota USA e por toda a empresa, e muitos começaram a adotar o BI. Por exemplo, o ex-gerente da TLS, que agora dirige a Divisão de Serviços ao Cliente Toyota, usa painéis em seu escritório, como o fazem os CFOs e outros grandes executivos pela Toyota (p. ex., para gerenciar melhor as despesas, compras, etc.). Agora está claro que quanto mais pessoas usarem ferramentas de análise de dados, mais dinheiro a Toyota pode ganhar. O sistema foi atualizado de 2003 para 2005 e ferramentas são adicionadas continuamente, conforme necessárias. Esses e muitos outros benefícios do sistema ajudaram a empresa-mãe, Toyota Motor Corp., a alcançar as mais altas margens de lucro no mercado automotivo desde 2003. Além disso, as participações no mercado estão aumentando consistentemente. A propósito, a Toyota, sendo uma empresa ágil, começará a produzir robôs que ajudam o cliente (p. ex., robôs de serviço para os idosos) no ano 2010, acreditando que nessa área esteja uma grande oportunidade.

Finalmente, um estudo independente realizado pela IDC Inc. acerca da justificação do gerenciamento de desempenho de negócios (veja o Capítulo 5) e sistemas de BI indica que a Toyota alcançou um retorno de 506% sobre o investimento em BI. (O retorno de investimento médio para as 43 outras empresas citadas pela *Fortune 500* que participaram do estudo foi de 112%.)

Fontes: Compilado de D. Briody, “Toyota’s Business Intelligence: Oh! What a Feeling!”, *CIO Insight*, October 2005, um documento da Hyperion Solution Corporation (hyperion.com/customers/stories/us_toyota_motor.cfm),¹ and toyota.com (acessado em março de 2006).

Questões sobre a vinheta de abertura

1. De que maneiras o antigo sistema de informações criou problemas para a Toyota?
2. Que necessidades de informação dos gerentes são satisfeitas pelo novo sistema? Que decisões têm o apoio do sistema de BI?
3. Relacione o problema com a cadeia de fornecimento (das fábricas para os revendedores e clientes).
4. Liste as ferramentas de suporte a decisões citadas aqui.
5. Que vantagem estratégica a Toyota pode extrair desse sistema?
6. Relacione as decisões da Toyota de fazer robôs que auxiliam o cliente ao ambiente de constante mudança dos negócios.

¹ As URLs de websites são dinâmicas. Com a publicação desta obra, verificamos se os *websites* citados estavam válidos e ativos. Os sites que referimos no texto algumas vezes estão diferentes ou fora do ar porque as empresas mudam de nome, são compradas ou vendidas, sofrem fusão ou falem. Às vezes *websites* estão fora do ar para manutenção, reparo ou *redesign*. A maioria das empresas deixaram de usar a designação inicial “www” para seus sites, mas algumas ainda a utilizam. Se você tiver problemas para conectar em algum site mencionado aqui, por favor, tente acessá-lo novamente por meio de uma busca na Web. Geralmente, o novo site pode ser facilmente encontrado. Pedimos desculpas pelo transtorno.

O que podemos aprender com esta vinheta?

Esta vinheta ilustra um caso típico no qual o fluxo de informações não pode arcar com as necessidades dos gerentes: as informações são atrasadas, às vezes, imprecisas e não são compartilhadas por todos. Os antigos sistemas não suprem a necessidade de tomar decisões rápidas, avaliar grandes quantidades de informação que podem ser armazenadas em diferentes locais e colaborar. A solução é uma tecnologia chamada *business intelligence* que é baseada em um *data warehouse* e fornece uma vantagem estratégica. O principal objetivo deste livro é mostrar como isso é feito. Neste capítulo, fornecemos uma introdução ao campo de BI.

1.2

ALTERANDO OS AMBIENTES DE NEGÓCIOS E O SUPORTE COMPUTADORIZADO À DECISÃO

A vinheta de abertura ilustra como uma empresa global se destaca no mercado automotivo altamente competitivo. A Toyota, conhecida por suas técnicas pioneiras de gerenciamento manual, como *just in time* (JIT), passou de forma radical ao suporte computadorizado de suas operações. Para entender por que a Toyota e muitas outras empresas estão adotando o suporte computadorizado à decisão, que inclui o *business intelligence*, desenvolvemos um modelo intitulado *pressões de negócios, reações organizacionais e suporte computadorizado*, exibido na Figura 1.1.

O modelo de pressões – reações – suporte de negócios

O modelo de pressões – reações – suporte de negócios, como já indica o nome, é composto por três componentes, como mostrado na Figura 1.1: as pressões de negócios – que resultam da atual atmosfera de negócios; as ações (reações) – que as empresas tomam para responder a essas pressões (ou para tirar proveito das oportunidades disponíveis no ambiente); e o suporte computadorizado – que facilita o monitoramento do ambiente e aprimora as ações de resposta realizadas pelas organizações.

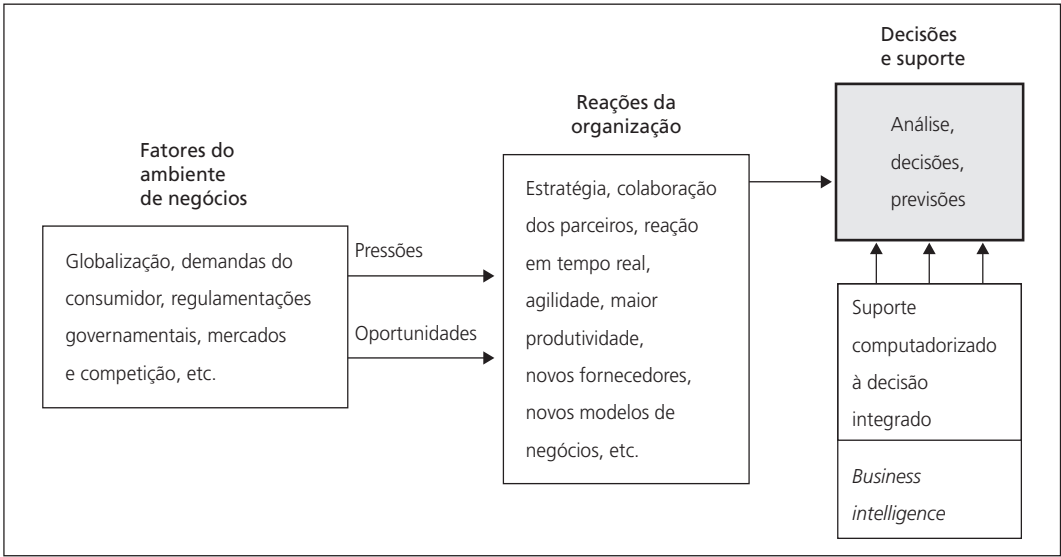


Figura 1.1 Modelo de pressões – reações – suporte de negócios.

O ambiente de negócios

O ambiente em que as organizações operam hoje em dia torna-se cada vez mais complexo (Huber: 2003). Por um lado, tal complexidade cria oportunidades; por outro, cria problemas. Tome como exemplo a globalização. Atualmente, é possível encontrar fornecedores e clientes em muitos países com facilidade, o que significa que se pode comprar material mais barato e vender mais seus produtos e serviços: existem grandes oportunidades. Porém, a globalização implica uma concorrência maior e mais forte. Dividimos os fatores do ambiente de negócios em quatro grandes categorias: *mercados, demandas do consumidor, tecnologia e sociedade*. Elas estão resumidas na Tabela 1.1.

Observe que a *intensidade* da maioria desses fatores aumenta com o tempo, gerando mais pressões, maior concorrência e assim por diante. Além disso, as organizações, ou certos departamentos dentro das organizações, lidam com orçamentos menores e pressão redobrada da alta administração para aumentar o desempenho e o lucro. Neste tipo de ambiente, os gerentes devem demonstrar reações rápidas, inovação e agilidade. Vejamos como fazer isso.

Reações organizacionais: ser reativo, precavido, versátil e proativo

As organizações privadas ou públicas têm consciência do atual ambiente de negócios e suas pressões. Elas usam diferentes ações para responder às pressões. A TLS da Toyota, por exemplo, usa o BI para aprimorar a comunicação e dar apoio aos executivos em sua tarefa de saber exatamente o que acontece em cada área de operação, quase em tempo real. Ela também usa o sistema para ter uma melhor colaboração com seus parceiros comerciais. Através disso, pode cortar gastos e melhorar a satisfação do cliente. Outras ações realizadas pelos gerentes podem incluir:

- Emprego de planejamento estratégico
- Uso de modelos de negócios novos e inovadores
- Reestruturação de processos de negócios
- Participação em alianças de negócios
- Aprimoramento dos sistemas de informações corporativos
- Aprimoramento das relações de parceria

Tabela 1.1 Fatores do ambiente de negócios que criam pressão nas organizações

<i>Fator</i>	<i>Descrição</i>
Mercados	Forte concorrência A expansão dos mercados globais A prosperidade dos mercados eletrônicos na Internet Métodos inovadores de <i>marketing</i> Oportunidades de terceirização com suporte de TI Necessidade de transações em tempo real e sob demanda
Demandas do consumidor	Deseja customização Deseja qualidade e diversidade de produtos e rapidez na entrega Os clientes tornam-se poderosos e menos fiéis
Tecnologia	Mais inovações, novos produtos e serviços A crescente taxa de obsolescência A crescente sobrecarga de informações
Sociedade	O aumento das regulamentações e desregulamentações governamentais Mão-de-obra diversificada conta com mais mulheres e de mais idade Segurança nacional e ataques terroristas são grandes preocupações Lei Sarbanes-Oxley e outras legislações relativas a prestação de contas são uma necessidade O crescimento da responsabilidade social das empresas

- Incentivo à inovação e à criatividade
- Aprimoramento do serviço e relacionamentos com clientes
- Adoção do comércio eletrônico
- Adoção da manufatura sob encomenda e da produção e serviços sob demanda
- Uso de novas tecnologias de informação para melhorar a comunicação, o acesso aos dados (descoberta de informações) e a colaboração
- Rápida reação às ações dos concorrentes, como em precificação, promoções e novos produtos e serviços
- Automatização de muitas tarefas do pessoal administrativo
- Automatização de certos processos de decisão, principalmente os que envolvem clientes
- Aprimoramento da tomada de decisões

Muitas dessas ações, se não todas, exigem algum suporte computadorizado.

Suporte computadorizado

Freqüentemente, essas e outras atividades de reação tornam-se mais fáceis com o auxílio de sistemas computadorizados de suporte à decisão, como o BI na Toyota.

Diminuindo a diferença estratégica Um dos principais objetivos do BI é facilitar a diminuição da diferença entre o desempenho atual de uma organização e o desempenho desejado, expresso em sua missão, objetivos, metas e a estratégia para atingi-los. Para maiores detalhes, consulte Coveney et al. (2003) e a discussão do Capítulo 5.

Questões de revisão da Seção 1.2

1. Liste os componentes do modelo de pressões – reações – suporte e explique-os.
2. Liste os principais fatores presentes no atual ambiente de negócios.
3. Liste algumas das maiores atividades de reação realizadas pelas empresas.
4. Defina a diferença estratégica.

1.3

UMA ESTRUTURA PARA O BUSINESS INTELLIGENCE (BI)

Conceitos e tecnologias de suporte à decisão foram implementados de forma crescente durante 30 anos, com diferentes nomes, por muitos fornecedores que criaram uma diversidade de ferramentas e metodologias. Em 2006, os principais produtos e serviços comerciais apareceram sob o nome “guarda-chuva” *business intelligence*, conforme citado na vinheta de abertura da Toyota.

Características do sistema da Toyota

Como você deve lembrar, o sistema da Toyota foi motivado por:

- Excesso de dados sem direção e dados com muitos erros.
- Forte concorrência no setor automotivo, principalmente da Honda.
- Problemas de comunicação e coordenação na cadeia de fornecimento.
- Incapacidade dos departamentos de compartilhar dados de forma oportuna.
- Incapacidade da gerência de obter as informações necessárias para a tomada de decisões.

Os sistemas da Toyota incluíram:

- Um *data warehouse* com dados históricos.
- Ferramentas para realizar análise e manipulação de dados.
- Uma *interface* visual de usuário (um *dashboard*), especialmente para os principais gerentes.

Conforme indica o caso, o uso do BI resolveu os problemas da empresa e foi um sucesso estrondoso. Vamos explorar o BI, começando por sua definição, e ver por que é uma tecnologia de tanto sucesso.

Definições de BI

Business intelligence (BI) é um termo “guarda-chuva” que inclui arquiteturas, ferramentas, bancos de dados, aplicações e metodologias (consulte Raisinghani: 2004). É uma expressão livre de conteúdo, portanto, significa coisas diferentes para pessoas diferentes. Parte da confusão relacionada ao BI é causada pela enxurrada de acrônimos e palavras da moda associadas a ele e suas ferramentas (como *business performance management* – BPM). Os principais objetivos do BI são permitir o acesso interativo aos dados (às vezes, em tempo real), proporcionar a manipulação desses dados e fornecer aos gerentes e analistas de negócios a capacidade de realizar a análise adequada. Ao analisarem dados, situações e desempenhos históricos e atuais, os tomadores de decisão conseguem valiosos *insights* que podem servir como base para decisões melhores e mais informadas (veja Zaman: 2005). O processo do BI baseia-se na *transformação* de dados em informações, depois em decisões e finalmente em ações.

Breve histórico

O termo *BI* foi cunhado pelo Gartner Group em meados da década de 1990. Contudo, o conceito iniciou muito antes, com suas raízes nos sistemas de geração de relatórios SIG dos anos 1970. Durante esse período, os sistemas de geração de relatórios eram estáticos, bidimensionais e não possuíam recursos de análise. No início dos anos 1980, surgiu o conceito de *sistemas de informações executivas (EIS)*. Esse conceito expandiu o suporte computadorizado aos gerentes e executivos de nível superior. Alguns dos recursos introduzidos foram sistemas de geração de relatórios dinâmicos multidimensionais (*ad hoc* ou sob demanda), prognósticos e previsões, análise de tendências, detalhamento, acesso a status e fatores críticos de sucesso. Esses recursos apareceram em dezenas de produtos comerciais até o meio da década de 1990. Depois, os mesmos recursos e alguns recursos novos apareceram sob o nome BI. Hoje, se reconhece que todas as informações de que os executivos necessitam podem estar em um bom sistema de informações empresariais baseado em BI. Assim, o conceito original de Sistemas de Informações Executivas foi transformado em BI. Em 2005, os sistemas de BI começaram a incluir o recurso de *inteligência artificial*, bem como recursos poderosos de análise. A Figura 1.2 ilustra as várias ferramentas e técnicas que podem ser incluídas no BI e também a sua evolução. Essas ferramentas proporcionam os recursos do BI. Os mais sofisticados desses produtos incluem a maioria dos recursos, outros se especializam apenas em alguns. Para mais detalhes, consulte Zaman (2005) e Raisinghani (2004).

Origens e direcionadores do *business intelligence*

De onde vieram as abordagens modernas? Quais são suas raízes, e qual foi o impacto dessas raízes na maneira como as organizações gerenciam essas iniciativas hoje? Os investimentos atuais em tecnologia da informação são examinados cada vez mais a fundo quanto a seu impacto e potencial nos resultados financeiros.

As organizações estão sendo forçadas a captar, compreender e explorar seus dados para dar suporte à tomada de decisões, a fim de melhorar as operações de negócios. Por um lado, a legislação e a regulamentação (p. ex., a Lei Sarbanes-Oxley de 2002) exigiram que os líderes de negócios documentem seus processos de negócios e atestem a legitimidade das informações em que confiam e que repassam aos acionistas. Além disso, o tempo do ciclo dos negócios agora está extremamente apertado; por isso, a tomada de decisões melhor, mais rápida e informada é uma obrigação competitiva. Os gerentes precisam das *informações certas na hora certa* e no *lugar certo*. Este é o mantra das

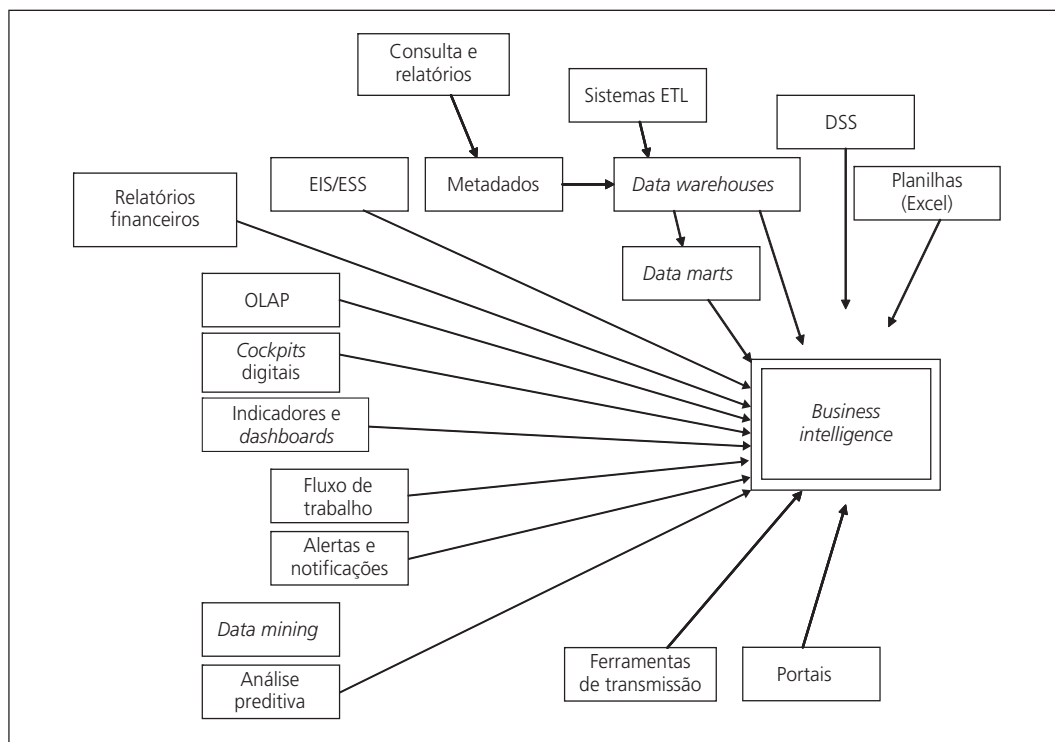


Figura 1.2 Evolução do BI.

abordagens modernas ao *business intelligence*. As organizações devem ser espertas. Prestar atenção especial ao gerenciamento das iniciativas de BI é um aspecto necessário ao se fazer negócios (Baum: 2006). Em vista disso, não é surpresa que as organizações estão promovendo o BI cada vez mais. A vinheta de abertura discutiu uma história de sucesso de BI na Toyota Motor Company. Você saberá mais sobre os sucessos de BI e seus fundamentos nos Capítulos 2 a 5. Exemplos de aplicações são fornecidos na coluna da esquerda na Tabela 1.2. Para maiores detalhes, consulte Thompson e Jakovljevic (2005), Biere (2003), Loschin (2003) e Baum (2006).

Arquitetura e componentes de BI

Conforme mostrado no caso da Toyota, o BI tem quatro grandes componentes: um *data warehouse* (DW) com seus dados-fonte a *análise de negócios*, uma coleção de ferramentas para manipular e analisar os dados no *data warehouse*, incluindo *data mining*; *business performance management* (BPM) para monitoria e análise do desempenho e uma *interface de usuário* (como o *dashboard*). Observe que o ambiente de *data warehousing* é sobretudo de responsabilidade de uma equipe técnica, e o ambiente de análise (também conhecido como *análise de negócios*) está no âmbito dos usuários de negócios. Qualquer usuário pode conectar-se ao sistema por meio de uma interface de usuário, como um navegador, e a alta administração pode usar o componente de BPM e também o *dashboard*. As ferramentas da análise de negócios e a interface de usuário serão introduzidas nos Capítulos 2-5. No entanto, não será descrito neste livro um conjunto de ferramentas, os *sistemas inteligentes*, que pode ser visto como um componente futurista do BI. Segundo Zaman (2005), este futuro desenvolvimento pode provocar a mudança de nome do campo para *inteligência artificial de negócios*.

Tabela 1.2 Valor para o negócio das aplicações analíticas de BI

<i>Aplicação analítica</i>	<i>Pergunta de negócios</i>	<i>Valor para o negócio</i>
Segmentação dos clientes	Em quais segmentos de mercado meus clientes se encaixam e quais são suas características?	Personalizar os relacionamentos com clientes para obter maior satisfação e retenção destes.
Propensão à compra	Quais clientes são mais propensos a responderem à minha promoção?	Visar os clientes com base nas necessidades que eles têm de aumentar a fidelidade à sua linha de produtos. Aumentar também a lucratividade da campanha enfocando quem tem mais possibilidades de comprar.
Lucratividade dos clientes	Qual é a lucratividade vitalícia de meu cliente?	Tomar decisões individuais de interação de negócios com base na lucratividade geral dos clientes.
Detecção de fraudes	Como posso saber quais transações provavelmente serão fraudulentas?	Determinar a fraude com rapidez e agir imediatamente para minimizar os custos.
Perda de clientes	Qual cliente está a ponto de sair?	Evitar a perda de clientes de grande valor e abandonar os de menor valor.
Otimização do canal	Qual é o melhor canal para chegar a meu cliente em cada segmento?	Interagir com os clientes com base nas preferências deles e em sua necessidade de gerenciar os custos.

Fonte: Figura 1 de A. Ziama e J. Kasher (2004), *Data Mining Primer for the Data Warehousing Professional*, Dayton OH:Teradata, uma divisão da NCR.

A arquitetura de alto nível do BI é apresentada na Figura 1.3. Ela é descrita nas subseções seguintes.

O data warehouse

Partindo do lado esquerdo da figura, vemos o fluxo de dados dos sistemas operacionais (p. ex. CRM, ERP etc.) até um DW, que é um banco ou repositório de dados especial preparado para dar suporte a aplicações de tomada de decisão. As aplicações variam de simples gerações de relatórios ou consultas a complexas otimizações. O DW é construído com as metodologias, principalmente metadados e ELT, descritas no Capítulo 2. Também ali descritos estão os *data marts*, repositórios de um assunto ou departamento específico (p. ex., *marketing*).

Análise de negócios

Há muitas ferramentas de *software* que permitem aos usuários criarem relatórios e consultas sob demanda e realizarem análises de dados. Elas surgiram originalmente com o nome de *processamento analítico online* (OLAP). Por exemplo, os usuários podem analisar diferentes dimensões de dados, como uma série temporal de vendas em cada região e a análise de tendências. Assim, os usuários de negócios são capazes de identificar com rapidez e facilidade as tendências de desempenho. Essa identificação é feita com a análise de informações cíclicas e recursos de gráficos de produtos que suportem análises de dados mais sofisticadas e integrem recursos completos de campos calculados aos relatórios. Por exemplo, os usuários podem rapidamente isolar e identificar produtos, clientes, regiões ou outras áreas que apresentam tendências significativas de alta ou de baixa ou que constituem fonte de problemas. Algumas soluções incluirão também uma função potente e totalmente integrada de gráficos de dados, que permite aos usuários criar visualizações de dados detalhadas. Esses recursos, além de outros mais so-

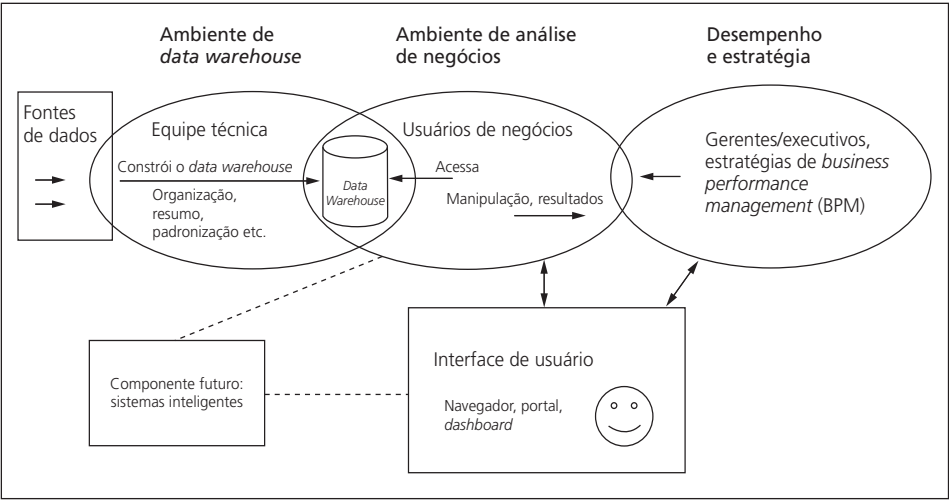


Figura 1.3 Uma arquitetura de alto nível do BI.

Fonte: desenho de E. Turban.

fisticados, são apresentados no Capítulo 3. Para realizar análises de negócios, o usuário precisa de um *software* de interatividade chamado *middleware* (para acessar o *data warehouse*). Ele é considerado *infra-estrutura*, assim como a *interface* de usuário ao sistema.

Um campo analítico emergente é o dos sistemas automatizados de decisão.

Sistemas automatizados de decisão. Uma abordagem relativamente nova ao suporte à tomada de decisão é conhecido como **sistemas automatizados de decisão (ADS)**, conhecido ainda como *sistemas de automatização de decisão (DAS)* (Davenport: 2006; Davenport E Harris: 2005). Eles são sistemas baseados em regras que normalmente oferecem uma solução em uma área funcional (como finanças ou produção) a um problema de gestão específico e repetitivo comum em um setor do negócio. Por exemplo, se aprovar ou não um pedido de empréstimo, ou, conforme ilustrado no Caso de aplicação 1.1, como determinar o preço de um item em uma loja.

CASO DE APLICAÇÃO 1.1

Determinação inteligente de preços usando o suporte automatizado à decisão

A precificação de alguns milhares de itens na Longs Drug Stores (uma cadeia norte-americana de cerca de 400 farmácias, **longs.com**) é descentralizada. Cada loja tem autonomia para determinar o preço dos itens que comercializa para oferecer melhor concorrência em termos locais. Normalmente, os preços eram atribuídos de forma manual, com a modificação do preço de varejo sugerido pelo fabricante. Práticas semelhantes existiam na maior parte das outras cadeias de varejo, inclusive su-

permercados. Além do mais, quando havia uma guerra de preços ou chegava uma determinada estação do ano, ocorria um corte geral de preços, sem que se desse atenção à previsão de demanda, rentabilidade, estratégia de precificação ou consistência de preços entre as lojas.

Agora, a precificação está passando por uma mudança radical, em grande parte como resultado do aprimoramento dos sistemas de suporte de TI. Indo na esteira do que fizeram linhas aéreas e locadoras de veí-

culos há anos, o setor varejista, incluindo a Longs Drug Stores e cerca de metade dos outros varejistas dos EUA, está introduzindo programas de *otimização de preços*. Como esses programas funcionam? Os programas de otimização de preços (p. ex. o oferecido pelo SAS Inc., **sas.com**, e outros) unem as regras de negócios a certos algoritmos de cálculo e formam um sistema que recomenda automaticamente um preço para cada item em cada loja. Os dados de entrada usados são os números de vendas da estação, as vendas reais de cada loja (em tempo real), a curva de preço-demanda de cada produto, os preços dos concorrentes, as métricas de rentabilidade, entre outros. Com o uso do programa, os varejistas podem identificar os produtos mais influenciados pelo preço e testar, em segundos, o impacto que uma alteração no preço poderia causar na margem de lucro (ou

outra meta desejada, como volume de vendas). Ao se utilizarem as prioridades, políticas e restrições de cada loja, é possível desenvolver e testar estratégias.

Os resultados na Long Drugs, e em outras lojas que usaram programas semelhantes, demonstram um aumento de volume, receita e perfil entre 2 e 10%. O *software* ainda é bastante caro; por isso, atualmente (2006) apenas grandes varejistas podem arcar com as despesas de utilização. Conforme mais concorrentes produzirem *software* similar, seu preço diminuirá e mais lojas o usarão. Os consumidores serão os maiores beneficiários, pagando menos pelos itens.

Fontes: compilado de A. Cortese, “The Power of Optimal Pricing”, *Business 2.0*, September 2002, e de **sas.com** (acessado em março de 2006).

Data mining

Data mining é uma classe de análise de informações, baseada em bancos de dados, a qual procura padrões ocultos em uma coleção de dados que podem ser usados para prever comportamentos futuros. Ele pode, por exemplo, ajudar as empresas de varejo a encontrar clientes com interesses em comum. Entretanto, o termo é comumente usado de forma equivocada para descrever *software* que apresentam dados de novas maneiras. O verdadeiro *software* de *data mining* não muda apenas a apresentação: ele de fato *descobre* relações antes desconhecidas entre os dados e este conhecimento é aplicado para se alcançar metas de negócios específicas. As ferramentas de *data mining* são usadas para substituir ou aprimorar a inteligência humana devido à sua capacidade de verificar enormes armazéns de dados. Desta forma, elas descobrem novas e significativas correlações, padrões e tendências através de tecnologias de reconhecimento de padrões e métodos estatísticos avançados. O *data mining* é descrito no Capítulo 4; as tecnologias que o alimentam, principalmente as redes neurais, são descritas no Capítulo online 6. No Tutorial online 2, “Amostra de aplicação do *software* Estatística”, está disponível uma ilustração prática de como um projeto de aplicação de rede neural é concluído usando um *software* comercial.

Business performance management

O componente final do processo de BI é *business (ou corporate) performance management* (BPM). Este componente baseia-se na metodologia *balanced scorecard*, que se trata de uma estrutura para definir, implementar e gerenciar a estratégia de negócios de uma empresa conectando objetivos a medidas factuais. Em outras palavras, é uma forma de conectar métricas de nível superior, como as informações financeiras criadas pelo *diretor financeiro* (CFO), a desempenhos reais de todos os níveis hierárquicos da corporação. O BPM usa a análise, a geração de relatórios e as consultas de BI. Seu objetivo é otimizar o desempenho geral de uma organização. Os detalhes do BPM são apresentados no Capítulo 5.

Atualmente, a maioria das suítes de BI permite a aplicação de *balanced scorecards* através da capacidade de comparar imediatamente o desempenho de negócios às metas estabelecidas. Elas também oferecem uma plataforma para compartilhamento de metas de desempenho e resultados em toda uma empresa, permitindo que a gerência rapidamente compreenda como vão os negócios.

O BPM normalmente inclui *dashboards*, que proporcionam uma visão rápida e abrangente do desempenho corporativo por meio de apresentações gráficas, bastante parecido com os instrumentos no painel de um carro. Essas apresentações gráficas exibem medidas, tendências e exceções de desempenho e integram informações de múltiplas áreas comerciais. O ponto central de qualquer projeto de *dashboard* são as métricas captadas e os indicadores de desempenho que são comparados ao desempenho real e combinados para formar gráficos que refletem a saúde da empresa.

Interface de usuário: dashboards e outras ferramentas de transmissão de informações

Os *dashboards* fornecem uma visão abrangente e visual das medidas (*indicadores-chave de desempenho*), tendências e exceções do desempenho corporativo provenientes de múltiplas áreas do negócio. Os gráficos mostram o desempenho real em comparação às métricas desejadas, propiciando uma visão imediata da saúde da organização. Outras ferramentas que “transmitem” informações são portais corporativos, *cockpits* digitais e outras ferramentas de visualização.

Ferramentas de visualização. Muitas ferramentas de visualização, desde apresentações em cubo multidimensional até a realidade virtual, são parte integral dos sistemas de BI. Lembre-se de que o BI surgiu do EIS; assim, muitos auxílios visuais aos executivos foram incorporados à *software* de BI.

Os benefícios do BI

Conforme foi visto no caso da Toyota, o principal benefício do BI para uma empresa é sua capacidade de fornecer informações precisas quando necessário, incluindo uma visão em tempo real do desempenho corporativo geral e de suas partes individuais. Tais informações são uma necessidade para todos os tipos de decisão, para o planejamento estratégico e mesmo para a sobrevivência. Eckerson (2003) revela os resultados de uma pesquisa entre 510 corporações que indica os benefícios do BI conforme a visão dos participantes. São eles:

- Economia de tempo (61%)
- Versão única da verdade (59%)
- Melhores estratégias e planos (57%)
- Melhores decisões táticas (56%)
- Processos mais eficientes (55%)
- Economia de custos (37%)

Thompson (2004) relatou, a partir de uma *survey*, que os maiores benefícios do BI são:

- Geração de relatórios mais rápida e precisa (81%)
- Melhor tomada de decisões (78%)
- Melhor serviço ao cliente (56%)
- Maior receita (49%)

Note que muitos dos benefícios são *intangíveis*; é por isso que, segundo Eckerson (2003), tantos executivos *não* insistem em uma justificativa de custos rigorosa para os projetos de BI. Thompson (2004) também apontou que as áreas mais comuns de aplicação do BI são relatórios gerais, análise de vendas e *marketing*, planejamento e previsão, consolidação financeira, relatórios regulamentares, orçamento e análise de rentabilidade. Uma aplicação interessante de *data mining* que usa ferramentas de *análise preditiva* é ilustrada no Caso de aplicação 1.2.

CASO DE APLICAÇÃO 1.2**Análise preditiva ajuda o Texas a arrecadar impostos**

Em muitas entidades públicas, há diferenças entre os impostos devidos e o total arrecadado. O estado do Texas, nos Estados Unidos, não é uma exceção. Para superar os problemas, os cobradores de impostos com frequência realizam auditorias que são dispendiosas e tomam tempo. Além disso, muitas auditorias são improdutivas, o que resulta em pouca ou nenhuma recuperação do imposto por ser difícil determinar quem será auditado. A fim de tomar melhores decisões relativas a auditoria e, assim, aumentar a porcentagem de auditorias produtivas, o Texas usa a *análise preditiva*.

Milhões de registros são armazenados no *data warehouse* do estado. Com o uso de *software* baseado em *data mining* da spss.com, milhões de registros podem ser cruzados para identificar orientações promissoras.

Especificamente, o sistema ajudou a identificar milhares de empresas que mantinham operações no estado sem cumprir com suas obrigações fiscais. Ele também ajudou os auditores fiscais a fazerem melhores seleções de alvos para auditoria. Uma vez que ganharam confiança no programa, os funcionários começaram a usá-lo amplamente, economizando mais de US\$ 150 milhões ao ano.

Fontes: compilado L. Gates, “State of Texas Recovers \$400 Million through Predictive Analytics”, *ADTmag.com*, May 26, 2005, adtmag.com/article.asp?id=11214 (acessado em março de 2006), e de Staff, “SPSS Predictive Analytics Helps Texas Recover \$400 Million in Unpaid Taxes”, *B-eye-network.com*, May 16, 2005, b-eye-network.com/view/868 (acessado em março de 2006).

O valor para o negócio do BI

As metodologias para avaliar o valor para o negócio do BI são complexas devido aos muitos benefícios intangíveis e ao DW ser uma infra-estrutura que serve a muitas aplicações. (Para ver as metodologias sugeridas, consulte Gartner: 2004 e Williams and Williams: 2003.) Uma forma de apresentar o valor para o negócio do BI é propor questões de negócios e mostrar como são respondidas com o suporte do BI, como apresenta a Tabela 1.2.

Como o BI pode ajudar

Gerentes e executivos necessitam de soluções de BI para gerenciar melhor seus negócios. As empresas que não conseguem implementar adequadamente essas soluções se colocam em uma situação de desvantagem competitiva. Para terem sucesso no ambiente de negócios de hoje, as empresas devem:

- Avaliar sua disponibilidade para enfrentar os desafios impostos por essas novas realidades do mercado.
- Ter uma abordagem holística da funcionalidade do BI.
- Fazer uso das práticas recomendadas e antecipar custos ocultos.

Gartner Inc. (2004) sugeriu as seguintes questões e estruturas cruciais para análise de BI:

- Como as empresas podem maximizar seus investimentos em BI?
- De que funcionalidade do BI as empresas precisam e o que estão usando hoje?
- Quais são alguns dos custos ocultos associados a iniciativas do BI?

Trataremos dessas questões nos Capítulos 2 a 5.

A seguir, um exemplo de como o BI ajudou uma empresa a lidar com a Lei Sarbanes-Oxley de 2002 (retirado de Gartner Inc., 2004 e Rogalski e Lin, 2003).

Exemplo 1

A Lei Sarbanes-Oxley (SOX) levou uma firma a implementar um novo sistema de gestão do desempenho financeiro capaz de atender às novas exigências de:

- Realizar análises e compilações impecáveis de milhares de transações e lançamentos.
- Equilibrar melhor o acesso aos dados, em oposição à necessidade de controlar o acesso a informações confidenciais internas.
- Fornecer relatórios à Comissão de Valores Mobiliários norte-americana em menos tempo.

A empresa implementou uma infra-estrutura e aplicações de BI que cumpriram os desafios. Dentro do propósito maior de obter a conformidade dos relatórios contábeis, os objetivos incluíram:

- Ficar "de olho" nos dados e nos indicadores-chave de desempenho e integrar controles rigorosos de segurança.
- Fornecer relatórios ativos que permitissem uma análise minuciosa até o nível mais inferior de detalhes da transação.
- Voltar a atenção para o tratamento contábil dos componentes materiais.
- Fazer uma busca proativa por anomalias nos bancos de dados financeiros usando *triggers* de variação.
- Reunir todos os dados financeiros em um banco de dados coeso.

Uma implementação intimamente ligada a esses objetivos deu à empresa um sistema de gestão do desempenho financeiro que possibilitou para a análise complementar aplicações de contabilidade e orçamento. Isso resultou em geração flexível de relatórios, investigação livre e análise de dados automatizada. A infra-estrutura e as aplicações de BI deram suporte a muitas quantidades e tipos de usuários e aplicações, como *data mining* automático para detecção de anomalias. Ela conseguiria alertar indivíduos específicos proativamente sempre que fosse detectada uma anomalia.

Questões de revisão da Seção 1.3

1. Defina BI.
2. Descreva o surgimento do BI.
3. Liste e descreva sucintamente os principais componentes do BI.
4. Liste os principais benefícios do BI.

1.4

CRIAÇÃO E USO DA INTELIGÊNCIA E GOVERNANÇA DO BI**Um processo cíclico de criação e uso da inteligência**

As iniciativas de *data warehousing* e BI normalmente seguem um processo semelhante àquele usado nas iniciativas de inteligência militar. Na verdade, os adeptos do BI seguem com frequência o modelo de segurança nacional ilustrado na Figura 1.4. O processo é cíclico com uma série de etapas inter-relacionadas. Uma das etapas mais importantes é a *análise*, ou seja, a conversão de dados brutos em informações que dão suporte à decisão. Entretanto, a análise não é precisa, confiável ou mesmo possível a menos que as outras etapas do caminho tenham sido abordadas de forma correta. O processo e suas etapas são detalhados no Arquivo *online* W1.1 e em Krizan (1999). Mais detalhes encontram-se a seguir.

Governança do BI

Uma vez que o *data warehouse* está estabelecido, o processo geral de *criação de inteligência* começa pela identificação e priorização de projetos *específicos* de BI (veja, por exemplo, Dignan: 2003). Em cada possível projeto de BI no portfólio, é importante usar as medidas de retorno do investimento (ROI) e custo total de propriedade (TCO) para estimar a relação custo/benefício. Isso quer dizer que, para cada projeto, devem ser examinados os custos associados às fases do processo em geral e também à manutenção da aplicação para o usuário de negócios. Além do mais, a estimativa de benefícios deve envolver análises do usuário final sobre o impacto nas decisões,

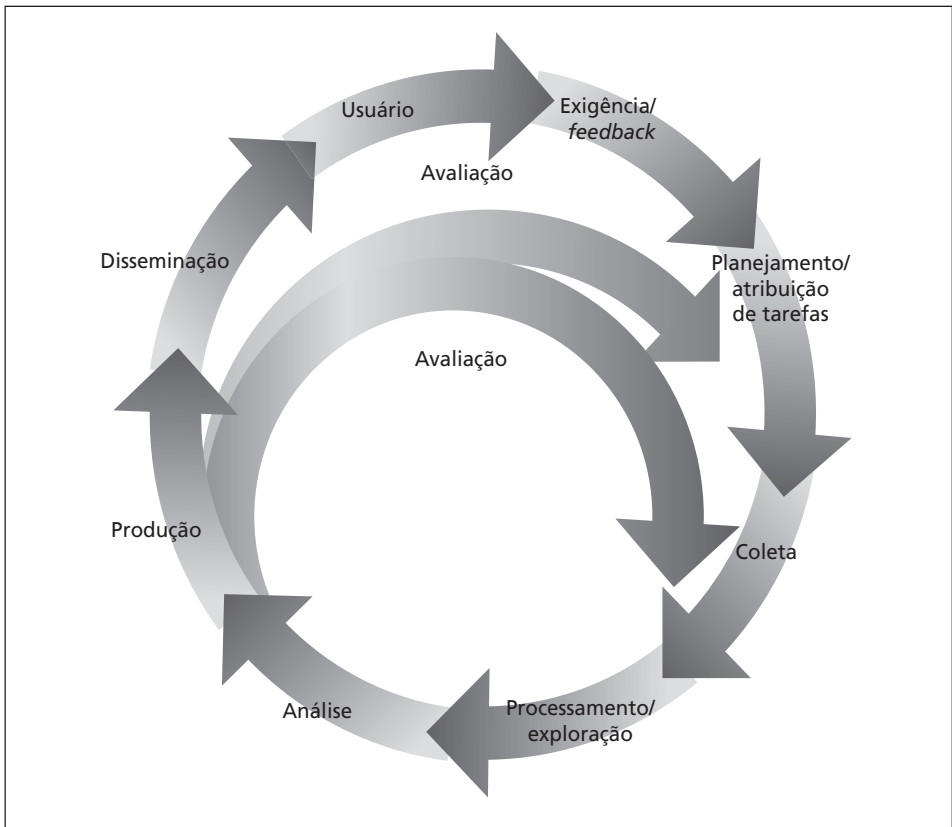


Figura 1.4 Processo de criação e uso da inteligência.

Fonte: L. Krizan, *Intelligence Essentials for Everyone*. Washington DC: Joint Military Intelligence College (occasional paper number six) Department of Defense, p. 6.

incluindo medidas de benefícios como a aceleração do fluxo de caixa. Algumas organizações se referem ao processo de priorização de projetos como uma forma de **governança do BI** (consulte Matney and Larson: 2004 e o Arquivo *online* W1.2).

Uma questão essencial de governança é quem deve atuar como tomador de decisão envolvido na priorização dos projetos de BI. Alguns sugerem (p. ex. Dignan: 2003) que a priorização bem-sucedida requer a “integração em nível gerencial”. Por exemplo, é possível caracterizar as pessoas pela função organizacional que exercem: estratégica, gerencial, operacional, clientes e fornecedores. As duas parcerias críticas exigidas para o controle do BI são: 1) uma parceria entre chefes de áreas funcionais e/ou líderes de áreas de produtos/serviços (gerencial); 2) uma parceria entre clientes e fornecedores em potencial (representantes do negócio e representantes da tecnologia da informação). Os gerentes podem lançar um olhar panorâmico sobre a organização e assegurar que as prioridades do projeto reflitam as necessidades de toda a empresa: eles podem garantir que um projeto não sirva apenas para subutilizar uma área em relação a outras. Os clientes podem oferecer idéias sobre a possível utilidade da inteligência gerada em um projeto, e os fornecedores são importantes à medida que refletem a realidade do abastecimento. Eis um conjunto típico de questões com as quais a equipe de controle do BI lidará: 1) criar categorias de projetos (estratégicos, obrigatórios, de investimento, de oportunidades de negócios, etc.); 2) definir critérios de seleção de projetos; 3) determinar e estabelecer uma estrutura para gerenciar o risco do projeto; 4) gerenciar e fazer uso das interdependências do projeto; 5) monitorar e ajustar continuamente a composição do portfólio.

Inteligência e espionagem

Embora muitos acreditem que o termo *inteligência* seja um sinônimo cheio de sigilo e mistério para operações clandestinas, no estilo CIA, dedicadas a roubar segredos corporativos, essa idéia não pode estar mais longe da verdade. É claro que esse tipo de *espionagem* existe, mas nosso interesse aqui é em como as empresas modernas se organizam ética e legalmente para colher o máximo de informações possíveis de seus clientes, ambiente de negócios, acionistas, processos de negócios, concorrentes e de outras fontes de informações potencialmente valiosas. No entanto, coletar dados é apenas o começo. Enormes quantidades de dados precisam ser catalogadas, rotuladas, analisadas, classificadas e filtradas. Esses dados também devem ser submetidos a diversas outras operações a fim de se tornarem úteis na tomada de decisão e melhoria dos resultados financeiros. A importância desses processos aumenta a cada dia, à medida que as companhias perseguem e acumulam mais e mais dados. Por exemplo, o agravamento do crescimento exponencial da quantidade de dados brutos compõe o surgimento da *identificação por radio frequência* (RFID). As aplicações de RFID provavelmente farão parte de uma categoria de aplicações empolgante e de rápido crescimento na próxima geração de especialistas em BI (consulte Heinrich: 2005). Unido a novas abordagens para a sintetização de informações a partir de fontes de texto por meio de “*text mining*” (Tutorial online T1) e a partir da rede por meio de *web mining* (Capítulo 4), além dos avanços no uso da rede para reunir inteligência (Giovinazzo: 2002), o advento da RFID sugere que as organizações estão prestes a entrar em uma nova era explosiva em BI para o suporte de decisões.

O BI adaptou um conjunto de nomenclaturas, sistemas e conceitos que o distinguem claramente de sua equivalente, a inteligência nacional e internacional voltada à espionagem. Apesar disso, há muitas analogias entre as duas: o fato de que grandes esforços devem ser empreendidos para se chegar a uma coleção de respeitáveis fontes de inteligência, o processamento dessa inteligência para obtenção de clareza e confiabilidade, a análise da inteligência bruta para produção de informações úteis e acionáveis, e o desenvolvimento de mecanismos para a disseminação adequada da informação aos usuários corretos.

Questões de revisão da Seção 1.4

1. Liste as etapas de criação e uso da inteligência.
2. O que é a governança do BI?
3. O que é reunir inteligência?

1.5

AS PRINCIPAIS TEORIAS E CARACTERÍSTICAS DO BUSINESS INTELLIGENCE

Para ilustrar as principais características do BI, primeiramente mostraremos o que ele não é: em outras palavras, processamento de transações.

Processamento de transações versus processamento analítico

Todos nós conhecemos bem os sistemas de informação que suportam nossas transações cotidianas: saques em caixas eletrônicos, depósitos no banco, verificações com *scanner* de caixas registradoras no mercado e assim por diante. Esses sistemas de *processamento de transações* com frequência estão envolvidos na condução de atualizações no que podemos chamar de *banco de dados operacional*. Por exemplo, em uma transação de saque em caixa eletrônico, o saldo bancário deve ser reduzido segundo o valor retirado; um depósito no banco aumenta a quantia na conta; uma compra no mercado provavelmente aparece no cálculo da loja sobre o total de vendas do dia e deve provocar uma redução no estoque de itens adquiridos da loja, etc. Esses **sistemas de processamento de transações online (OLTP)** lidam com os negócios rotineiros no andamento de uma empresa. Por sua vez, um *data warehouse* é um sistema distinto que fornece armazenamento

para os dados que serão utilizados na *análise*. A intenção é dar à gerência a capacidade de buscar informações sobre a empresa nos dados e usar sua análise para proporcionar suporte tático ou operacional às decisões. Isso fará com que, por exemplo, o pessoal de linha possa tomar resoluções mais rápidas e/ou informadas. Daremos uma definição mais técnica de DW no Capítulo 2, mas basta dizer que a intenção dos DWs é trabalhar com os dados informativos usados nos **sistemas de processamento analítico online (OLAP)**.

A maioria dos dados operacionais nos sistemas ERP – e em seus irmãos complementares, como a *gestão da cadeia de suprimentos* (SCM) ou o *gerenciamento de relacionamento com o cliente* (CRM) – está armazenada em um sistema OLTP, que é um tipo de sistema de computador que responde imediatamente às solicitações do usuário. Cada solicitação é considerada uma *transação*, ou seja, um registro computadorizado de um evento separado, como o recibo de um estoque ou um pedido do cliente. Em outras palavras, uma transação requer um conjunto de duas ou mais atualizações no banco de dados que devem ser realizadas sob o princípio do tudo ou nada.

O próprio projeto que torna um sistema OLTP eficiente no processamento de transações o torna ineficiente em relatórios, consultas e análises *ad hoc* ao usuário final. Na década de 1980, muitos usuários de negócios referiam-se a seus *mainframes* como “o buraco negro”, pois as informações entravam ali, mas nenhuma jamais saía. Todas as solicitações de relatórios tinham de ser programadas pela equipe de TI e apenas relatórios “pré-enlatados” podiam ser gerados com agendamento prévio. Consultas *ad hoc* em tempo real eram praticamente impossíveis. Embora os sistemas ERP dos anos 1990 baseados em cliente/servidor facilitassem, até certo ponto, a geração de relatórios, eles ainda estavam muito longe do que desejavam os usuários finais normais e não-técnicos para tarefas como a geração de relatórios operacionais, a análise interativa, etc. Com a intensão de resolver esses problemas, foram criadas as noções de DW e BI.

Algumas teorias sobre business intelligence

Ao longo dos anos, foram desenvolvidas várias teorias sobre BI. Algumas são descritas a seguir.

Fábrica e warehouse

O termo *warehouse* (armazém, em português) é associado ao conceito de uma fábrica. As fábricas têm seus próprios armazéns, recebem mercadorias dos armazéns e lhes entregam produtos acabados. E, de fato, usa-se a expressão “fábrica de informações da empresa” para descrever a forma como as empresas conduzem e organizam os esforços de *business intelligence* (Inmon: 2005).

A fábrica de informações

O conceito de fábrica de informações vê BI/DW como um componente central e crítico de qualquer organização (corporativa ou governamental) e se volta cada vez mais para o ambiente da Web. Como ocorre em uma fábrica de verdade, aqui também há *dados de entrada* (fontes e aquisição de dados), *armazenamento* (DW, *data marts*), *processamento dos dados de entrada* (análise, *data mining*) e *dados de saída* (entrega de dados, aplicações de BI). Detalhes sobre a fábrica de informações para empresas e governos estão disponíveis em **inmoncif.com**. A fábrica de informações é conectada a outros sistemas internos de informação – como ERP, CRM e comércio eletrônico – e também a sistemas externos de informação (normalmente, via *Internet* ou *extranet*). O conceito é ilustrado na Figura 1.5.

Data warehousing e business intelligence

Um DW é uma coleção de dados projetada para oferecer suporte à tomada de decisões gerenciais. *Data warehouses* contêm uma grande variedade de dados que apresentam uma imagem coerente das condições da empresa em um determinado ponto no tempo. A idéia por trás do conceito foi criar uma infra-estrutura de banco de dados que estivesse sempre *online* e contivesse todas as informa-

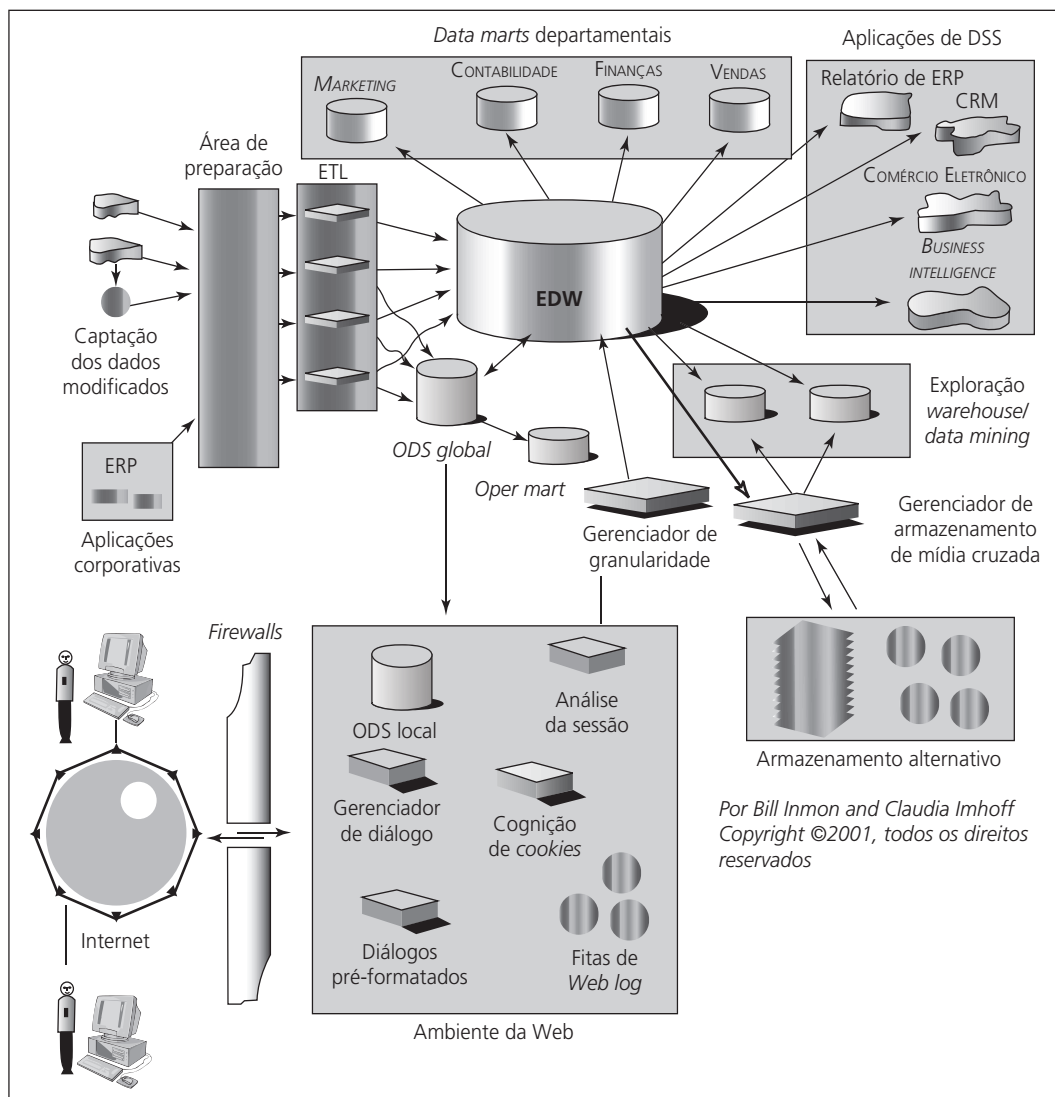


Figura 1.5 A fábrica de informações corporativa.

Fonte: Inmon Data Systems (inmoncif.com/library/cif)

ções dos sistemas OLTP, incluindo dados históricos. Porém, esta infra-estrutura seria reorganizada e estruturada de forma a oferecer rapidez e eficiência em consultas, análises e suporte à decisão.

Metodologia de análise avançada da Teradata

A Teradata, uma divisão da NCR, criou outra metodologia para a BI, exibida na Figura 1.6. Como mostra a figura, as aplicações de BI (lado superior esquerdo) são suportadas por técnicas e ferramentas de análise avançada (lado esquerdo). A metodologia é apresentada no lado direito como um processo cíclico cercando o *data warehouse* empresarial (EDW), que inclui etapas como compreensão dos negócios e compreensão dos dados (lado direito da figura). A metodologia oferece uma abordagem voltada aos negócios a qual abrange todas as técnicas que ajudam a construir mo-

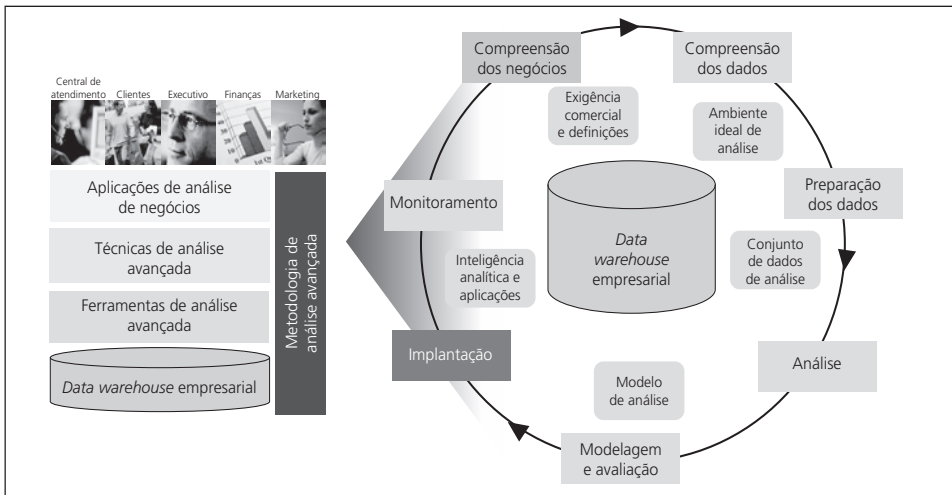


Figura 1.6 Metodologia de análise avançada da Teradata.

Fonte: cortesia da Teradata, uma divisão da NCR Corp.

delos, possibilitam novas visões dos dados, auxiliam na simulação para criar diferentes cenários, ajudam a entender as realidades e a prever futuros estados e resultados. (Há detalhes disponíveis nas diferentes edições da revista *online* gratuita *Teradata Magazine*, teradatamagazine.com.)

Separar o OLTP da análise e do suporte à decisão ativa os benefícios da BI descritos anteriormente, e permite inteligência e vantagem competitiva, como será descrito a seguir.

Estrutura de BI da Oracle

A Oracle Inc. (oracle.com) é conhecida por seu trabalho de integrar aplicações, bancos de dados e análise. A Figura 1.7. ilustra a estrutura conceitual de como a Oracle visualiza o papel do BI na empresa.

Questões de revisão da Seção 1.5

1. Defina processamento de transações *online* (OLTP).
2. Defina processamento analítico *online* (OLAP).
3. Descreva a fábrica de informações.
4. Descreva a estrutura de BI da Teradata.
5. Descreva a estrutura conceitual integrada de BI da Oracle.

1.6

RUMO À INTELIGÊNCIA E À VANTAGEM COMPETITIVAS

Muitos acreditam que a maior contribuição do BI é seu papel de dar às empresas uma vantagem estratégica.

A obrigação estratégica do BI

Os projetos de BI demonstraram ter valor agregado significativo para as organizações (Williams e Williams: 2003). Além do ROI e outros benefícios tangíveis, há cada vez mais evidências de que as iniciativas de BI se tornam uma *obrigação estratégica*. Alguns benefícios do BI estão determinando esta crença emergente. Primeiramente, em muitos mercados, as barreiras à entrada de um

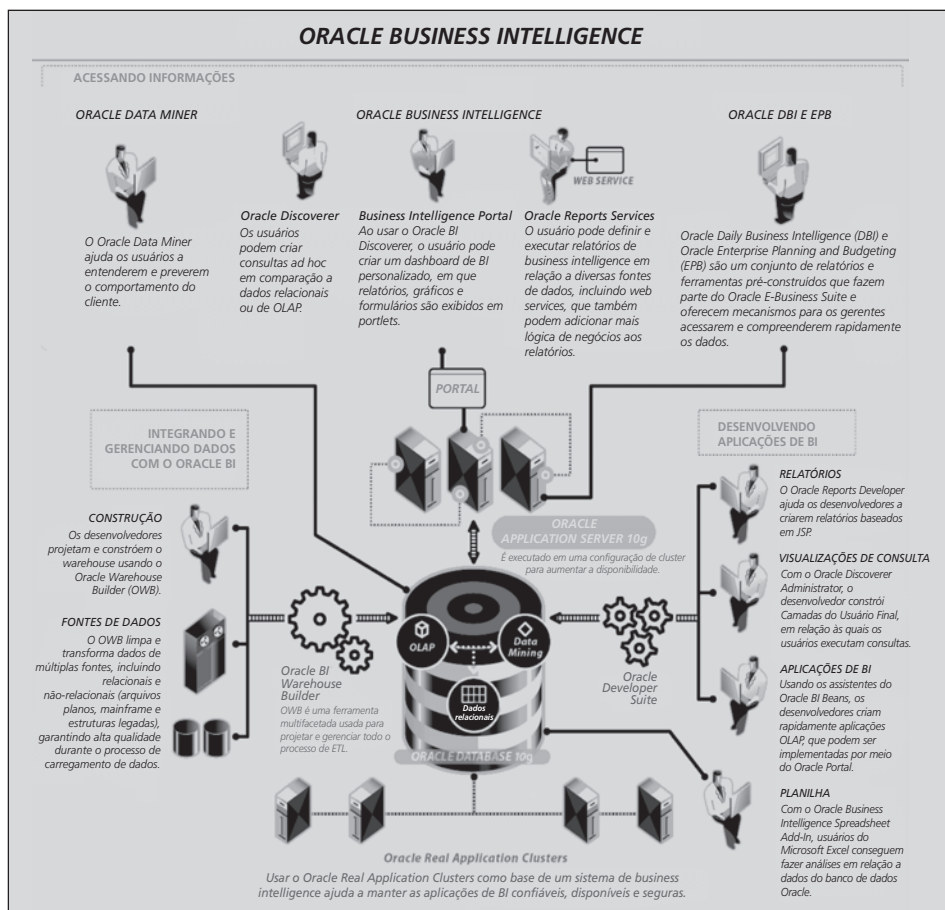


Figura 1.7 Sistema de BI da Oracle.

Fonte: A. Joch. "Eye on Information." *Oracle Magazine*, January/February 2005.

novo concorrente estão diminuindo de maneira significativa. Isso mostra que mesmo uma organização que ocupa uma posição sólida dentro do setor pode ter que enfrentar novos concorrentes, pois foram reduzidos os custos e outras restrições para que uma empresa se torne um protagonista no mercado. Devido à revolução da Web e à globalização cada vez mais intensa, por exemplo, empresas de todo o mundo estão desafiando grandes protagonistas de setores como fabricação de chips e computadores, produção automobilística, eletrônicos, têxtil e desenvolvimento de *software*, entre outros. Além disso, a capacidade de expedir mercadorias para todo o mundo por meio de transportadores prontamente disponíveis, como FedEx, UPS e DHL, bem como o comércio eletrônico, está facilitando que os possíveis concorrentes mandem seus produtos e serviços para mais clientes praticamente em qualquer lugar. Da mesma forma, as empresas encontram fornecedores melhores ou mais baratos em todo o planeta. As organizações hoje devem entender com atenção o que acontece em cada aspecto de suas operações, seu setor e seu ambiente de negócios e promover a capacidade de melhorar constantemente os processos quando necessário. Elas devem alterar ou adiantar a estratégia a fim de permanecerem um passo à frente das crescentes pressões competitivas, ou seja, devem ser versáteis e ágeis. Essas exigências estão relacionadas a um campo de estudo emergente chamado *inteligência competitiva* (CI).

Inteligência competitiva

Ainda que a CI com frequência envolva iniciativas de BI utilizadas na maioria das organizações, ela possui suas próprias ferramentas. Uma diferença entre as duas é que a CI implica em acompanhar o que os concorrentes estão fazendo; para isso, se reúne material sobre suas atividades recentes e em progresso. Por exemplo, informações competitivas importantes podem estar contidas em fontes incomuns, como transcrições de audiências judiciais, encontradas em documentos públicos de tribunais. Às vezes, informações relevantes são encontradas nas provas documentais de órgãos governamentais (p. ex. solicitações de patente). Em iniciativas de BI, algumas fontes de dados externas são incluídas no processo de análise, mas não raro esses dados estão disponíveis junto a fornecedores terceirizados. Por exemplo, pode ser que uma organização queira unir informações demográficas (como características da população, clima, etc.) a dados sobre os clientes, e compre esses dados demográficos para integrá-los a seus dados de clientes e fazer uma análise. Membros da Society of Competitive Intelligence Professionals (scip.org) vêem a *business intelligence* como um aspecto emergente de seu estatuto geral, que normalmente inclui análise competitiva mais genérica. Para maiores detalhes, consulte o Capítulo 3.

Estratégia competitiva em um setor

A análise dos concorrentes também é um componente da *análise do setor* que serve como base para o planejamento estratégico. Há várias estratégias genéricas comuns para analisar o setor. Por exemplo, uma organização pode optar por ser *líder de baixo custo* em um determinado mercado. Se ela conseguir manter sua estratégia de liderança de baixo custo, as barreiras à entrada serão altas, pois outra organização acharia difícil competir. A inteligência competitiva, neste contexto, implica que as empresas precisam saber se novas formas de produzir/gerar/fornecer um produto ou serviço resultariam em custos significativamente mais baixos, alterando, assim, a paisagem competitiva. As aplicações de BI aqui podem incluir exames minuciosos das métricas de qualidade associadas a processos de produção específicos, análise de matéria prima de vários fornecedores para avaliar taxas de defeitos, acompanhamento dos custos dos bens vendidos como porcentagem de volume produzido, etc. Além disso, as aplicações de BI geram regras de negócios que podem realmente ser integradas aos processos de negócios. Por exemplo, há uma forma de BI chamada *análise de clickstream* que usa informações reunidas de usuários que acessam um *website* (veja o Capítulo 3).

Outra estratégia competitiva é concentrar-se em um *nicho de mercado* em particular, talvez através de uma forma de diferenciação do produto ou serviço. Isso quer dizer que um segmento específico de mercado que tem uma preferência especial, talvez por produtos e serviços sofisticados, seria o alvo da estratégia. Há muitas aplicações de BI neste contexto, por exemplo, para se ter certeza de que as necessidades do cliente são satisfeitas e que há criação de fidelidade. Para isso, é importante acompanhar as preferências do cliente e lembrar dessas preferências no próximo encontro. Preparar um quarto de hotel segundo as preferências comuns do viajante de negócios e entregar-lhe o jornal desejado pela manhã pode ser uma forma de criar fidelidade neste lucrativo nicho de mercado. Uma empresa de jogos de azar como a Harrah's (consulte o Exercício 4 deste capítulo e Evans: 2006) pode adotar uma abordagem para a criação de fidelidade baseada em BI. Essa abordagem é dar aos clientes o mesmo tratamento em qualquer um de seus estabelecimentos em lugares diferentes, e disponibilizar informações sobre o programa de fidelidade relacionado em cada local. As recompensas dadas aos clientes quando visitam um local da Harrah's se acumulam, criando um cliente fiel e que retorna sempre: este é o nicho de mercado de alto valor visado pela Harrah's. Além disso, um programa de recompensas como esse cria uma *diferenciação* de produtos e serviços, em comparação aos cassinos maiores e mais luxuosos que podem existir em apenas um lugar.

Mantendo a vantagem competitiva

Como acontece com as outras iniciativas de TI, tentar obter uma vantagem competitiva usando o BI é apenas um objetivo. Mais importante é a capacidade de *manter* esta vantagem. As organizações fazem isso através da criação de marca e fidelidade do cliente usando aplicações de BI que, como vimos, dão suporte a estratégias de diferenciação de produtos/nichos de mercado. A maioria dos analistas de estratégia concorda que a liderança de baixo custo pode não render uma vantagem a menos que esta seja mantida. Por esse motivo, os projetos de BI e os DW se tornam ingredientes cada vez mais importantes na manutenção da vantagem competitiva entre as empresas. O tipo de projeto de BI pode variar com base na estratégia, e a equipe de controle do BI, em especial, pode priorizar possíveis projetos com base em sua capacidade de manter a vantagem competitiva. Assim, mesmo com cálculos favoráveis de custo/benefício e de retorno do investimento, um aspecto intangível como obrigação estratégica e sua sustentabilidade é um dos principais motivos pelos quais as organizações adotam o BI.

Questões de revisão da Seção 1.6

1. Por que o BI tem uma obrigação estratégica?
2. O que é inteligência competitiva?
3. Como o BI pode oferecer uma vantagem competitiva em um setor?
4. Como o BI pode ajudar a manter a vantagem competitiva?

1.7

A IMPLEMENTAÇÃO BEM-SUCEDIDA DO BUSINESS INTELLIGENCE

Implementar e implantar uma iniciativa de BI pode ser uma operação vagarosa, cara e passível de falha. Vamos explorar algumas das questões envolvidas.

A comunidade típica de usuários de BI

A comunidade de usuários de BI é grande e diversificada. O sucesso do BI depende, em parte, de quais pessoas na organização mais provavelmente fariam uso dele. Um dos aspectos mais importantes de um sistema bem-sucedido de BI é que ele deve ser vantajoso para a empresa como um todo. Isso implica em inúmeros usuários na empresa, muitos dos quais devem estar envolvidos desde o início de uma decisão pelo investimento em DW. É de se esperar que provavelmente haja usuários concentrados no nível estratégico e outros mais voltados ao nível tático. Uma estrutura adequada para descrever as comunidades de usuários é uma discussão das seguintes categorias: fazendeiros, turistas, operadores, exploradores e mineradores (conforme sugestão de Imhoff e Pettit: 2004). Os detalhes que estão no Arquivo *online* W1.3. Gartner Inc. (2004) diferenciam seis tipos parecidos de usuários. A Tabela 1.3 mostra esses diferentes usuários, quantos existem (em uma empresa grande), que ferramentas de BI eles usam e o valor estratégico de seu uso.

As várias classes de usuários de BI em uma organização podem ajudar a orientar a estruturação do DW e os tipos de ferramentas de BI e outros *software* de suporte necessários. Os integrantes de cada grupo são uma fonte excelente de informações sobre como avaliar os custos e benefícios de projetos específicos de BI, quando há um DW em funcionamento. A partir da discussão acima, fica evidente que uma característica importante da empresa que tem sucesso em sua abordagem ao BI é a *valorização* adequada das *diferentes classes* de usuários em potencial.

Planejamento adequado e alinhamento com a estratégia de negócios

Acima de tudo, as razões fundamentais para se investir em BI devem estar alinhadas com a estratégia de negócios da empresa. O BI não pode ser simplesmente um exercício técnico para o departamento de sistemas de informação: ele deve servir como uma forma de mudar a maneira como a empresa conduz suas operações de negócio. Isso é feito através da melhoria dos processos de negócios e da

Tabela 1.3 Correspondência entre tipos de usuários e funcionalidades para obtenção do valor máximo

<i>Tipos de usuários</i>	<i>Equipe de TI</i>	<i>Usuários avançados</i>	<i>Executivos</i>	<i>Gerentes funcionais</i>	<i>Clientes de informação esporádicos</i>	<i>Extranet: parceiros, consumidores</i>
Número de usuários	Poucos	Dezenas	Dezenas	Dezenas a centenas	Centenas a milhares	Centenas a milhares
Ferramentas e funções de BI	Desenvolvedor Admin. Metadados Dados de segurança Gerenciamento Aplicações Integração	Consultas <i>ad hoc</i> Relatórios de OLAP <i>Data mining</i> Análise avançada	<i>Dashboard</i> Indicador Relatórios de CPM (<i>corporate performance management</i>) BPM	Relatórios Planilha Visão de OLAP BAM (monitoramento de atividades de negócios) CPM	Relatórios Planilha Consultas	Relatórios Acompanhamento
Valor estratégico	Baixo	Alto	Muito alto	Médio	Baixo	Alto

Fontes: compilado de Gartner Inc. (2004); Imhoff e Pettit (2004).

transformação das tomadas de decisão em processos mais orientados aos dados. Muitos consultores e adeptos de BI envolvidos em iniciativas de sucesso informam que ter uma estrutura para o planejamento é uma pré-condição necessária. Uma estrutura desenvolvida pelo Gartner Inc. (2004) decompõe o planejamento e a execução em componentes de *negócios*, *organização*, *funcionalidade* e *infra-estrutura*. Nos níveis de organização e de negócios, devem ser definidos objetivos estratégicos e operacionais enquanto se consideram as habilidades disponíveis na organização para se chegar a esses objetivos. A alta administração deve considerar as questões de cultura organizacional em torno das iniciativas de BI e a formação de entusiasmo para essas iniciativas e procedimentos. Para que se compartilhem as melhores práticas de BI dentro da organização; deve também haver planos para preparar a organização para as mudanças. Uma das primeiras etapas desse processo é avaliar a organização do sistema de informação, os conjuntos de habilidades das possíveis classes de usuários e se sua cultura é receptiva a mudanças. A partir dessa avaliação, e supondo que haja justificativa e necessidade de seguir adiante, a empresa pode preparar um plano de ação detalhado (consulte o Arquivo *online* W1.4 para mais detalhes). Outra questão crítica relativa à implementação bem-sucedida da BI é a integração de diversos projetos de BI (a maioria das empresas usa diversos projetos) entre si e com outros sistemas de TI na empresa e também com os parceiros de negócio.

Gartner Inc. (2004) preparou um relatório abrangente sobre a implementação do BI e suas relações com outros sistemas empresariais, como ERP e CRM (o relatório também apresenta estudos de caso interessantes). Os principais tópicos que o relatório cobre são:

- Tendências e tecnologias de *business intelligence*: a busca pelo *insight*.
- Abordagens eficientes de *business intelligence* para o atual mundo dos negócios.
- Organização para o sucesso do *business intelligence*.
- Melhores práticas para definir métricas eficientes de negócios.
- Construção de uma infra-estrutura ágil de inteligência estratégica de negócios.
- Os benefícios da gestão eficaz de qualidade dos dados e metadados.
- *Data warehousing* e *business intelligence*: gerenciar custos e agregar valor.
- Gestão do desempenho corporativo: tendências de negócios e melhores práticas.
- O roteiro de gestão do desempenho corporativo.
- Tendências-chave de controle corporativo e gestão de conformidade.
- Uso do monitoramento das atividades do negócio para obter vantagem em tempo real.

- Como tirar proveito máximo do ERP por meio da *business intelligence*.
- O papel da análise nas estratégias bem-sucedidas de CRM.
- Análise da *Web*: do *software* ao modelo de serviço.
- Obtenção da produtividade no trabalho com portais e suítes empresariais.

Se a estratégia da empresa estiver devidamente alinhada com as razões das iniciativas de DW e BI, e se a organização de seu sistema de informação for capaz de fazer seu papel no projeto, e ainda se a comunidade de usuários necessária estiver disponível e com a motivação certa, é uma sábia decisão adotar o BI e estabelecer na empresa um centro de competência de *business intelligence* (BICC). O que o BICC de uma empresa pode realizar? A seguir há algumas possíveis realizações, conforme exemplificado pela France Telecom (Caso de aplicação 1.3).

- O centro pode demonstrar como o BI está claramente vinculado à estratégia e sua execução.
- O centro pode servir para incentivar a interação entre as possíveis comunidades de usuários de negócios e a área de sistema de informação.
- O centro pode servir como repositório e disseminador das melhores práticas de BI entre as diferentes linhas de negócios.
- Padrões de excelência nas práticas de BI podem ser defendidos e incentivados por toda a empresa.
- A área de sistema de informação pode aprender muito através da interação com as comunidades de usuários, por exemplo, sobre a variedade de tipos necessários de ferramentas de análise.
- A comunidade de usuários de negócios e a área de sistema de informação podem entender melhor por que a plataforma de *data warehouse* deve ser flexível o bastante para admitir exigências comerciais que mudam constantemente.
- Ele pode ajudar acionistas consideráveis, como executivos de alto nível, a ver como a BI tem um papel importante.

Para mais informações sobre BICC, consulte Gartner Inc. (2004).

CASO DE APLICAÇÃO 1.3

O *business intelligence* da France Telecom

Em um curto período de tempo, a France Telecom passou de única provedora de telecomunicações da França, em um setor que estava desregulamentado, para uma de muitas. Com novos concorrentes entrando rapidamente no mercado de telecomunicações, os executivos da France Telecom sabiam que tinham de usar seus sistemas de informação como uma importante arma estratégica. Eles deram início a um reposicionamento com o lema de tornar-se uma “empresa de rede”. Isso exigiu uma reconsideração intensa a fim de promover a padronização dos sistemas de informação por toda a empresa e suas subsidiárias. Para tanto, os executivos começaram a migrar todas as aplicações para uma arquitetura técnica mais adequada aos recursos baseados na Web. No passado, a empresa era organizada regionalmente, e cada

unidade comercial regional gerenciava seu próprio orçamento de TI. Isso gerou a coexistência de uma infinidade de tecnologias, versões de *software* e outros elementos desiguais.

A padronização foi a primeira etapa do estabelecimento de uma base infra-estrutural para uma grande iniciativa de *data warehouse* e BI. A empresa instituiu uma equipe de quatro pessoas para facilitar o “centro de competência de *business intelligence*” (BICC). O BICC foi encarregado de fiscalizar a implementação do *data warehouse*, garantindo que diferentes unidades comerciais e equipes de BI compartilhassem melhores práticas, e mantendo a consistência em todos os projetos de BI. Algumas tarefas importantes foram de responsabilidade do BICC. Primeiro, ele foi incumbido de providenciar serviços

de consultoria e desenvolvimento, incluindo oferecer aos gerentes de projeto assistência sobre estratégias relativas a projeto, auditoria, instalação etc. Segundo, o BICC ofereceu suporte a gerentes de projeto, arquitetos, projetistas, desenvolvedores e operadores por uma linha direta e um *helpdesk*. O suporte também incluiu um *website* na *intranet* para dar consultoria e dicas de ferramentas de BI, consultoria em metodologia e documentação de instalação. Em seguida, o BICC foi designado negociador da empresa junto aos fornecedores de BI. O BICC centralizou a abertura de todos os arquivos de caso; pedidos de atualizações de produtos; certificações de novas versões de *software*; e o acompanhamento, distribuição e manutenção dos acordos de licença. Por fim, o BICC ajudou no suporte aos usuários finais com ferramentas que os tornassem mais autônomos, incluindo um site na *intranet* dedicado ao suporte ao usuário, treinamento online e ajuda interativa. O site na *intranet* também foi usado para divulgar informações de casos de sucesso de BI, isto é, projetos que resultaram em desempenho superior, alinhados com a estratégia e os objetivos da empresa. Com esse intuito, para os executivos o site foi uma fonte de informações sobre a situação da iniciativa de BI e como ela oferece valor comercial.

Em resumo, o BICC da France Telecom foi desenvolvido para ajudar a organização a administrar seu portfólio de projetos de BI, padronizar abordagens analíticas em toda a empresa, treinar e educar os usuários finais, ajudar os usuários avançados, oferecer gestão do conhecimento através de compartilhamento das melhores práticas, e cuidar de todas as relações e suporte ao fornecedor. O serviço da France Telecom a mais de 91 milhões de clientes em 220 países dos cinco continentes teve uma melhoria significativa por meio de seu DW e seus projetos de BI em andamento. Como exemplo do apoio da liderança executiva ao BICC e suas iniciativas, o diretor de operações dos sistemas de informação de relações com o cliente declarou: “para conquistar novos clientes e criar fidelidade, agora baseamos nossas ações em um processo de *business intelligence*, no qual a BI tem um papel-chave de recuperar e analisar dados sobre nossos recursos corporativos. Atualmente, a empresa tem 130.000 computadores, e em quase metade deles há *software* de BI”.

Fontes: compilado de Business Objects’ “France Telecom BI in Action”, *Business Objects’ Case Study*, 2002, businessobjects.com/pdf/success/France_Telecom.pdf (acessado em abril de 2006), e de Gartner Inc., *Using Business Intelligence to Gain a Competitive Edge*. Relatório especial, Stamford CT, 2004.

Para saber mais sobre BICCs, visite intelligentsolutions.com e dmreview.com.

Outro fator importante do sucesso da BI é sua capacidade de promover um ambiente ágil, em tempo real e sob demanda.

BI em tempo real e sob demanda é possível

A exigência por acesso instantâneo e sob demanda a informações dispersas cresceu, pois a necessidade de diminuir a distância entre os dados operacionais e os objetivos estratégicos tornou-se mais premente. Como resultado disso, surgiu uma categoria de produtos chamados aplicações de *BI em tempo real* (consulte o Capítulo 3). A introdução de novas tecnologias de geração de dados, como a identificação por radio frequência (RFID), apenas acelera este crescimento e a necessidade subsequente por BI em tempo real. Sistemas tradicionais de BI usam um grande volume de dados *estáticos* extraídos, limpos e carregados em um *data warehouse* para produzirem relatórios e análises. No entanto, a necessidade não é apenas de gerar relatórios, já que os usuários também precisam de monitoramento de negócios, análise de desempenho e de um entendimento do porquê as coisas acontecem. Então, aos usuários que precisam saber (praticamente em tempo real) sobre mudanças nos dados ou disponibilidade de relatórios relevantes, podem ser oferecidos alertas e notificações sobre eventos e tendências emergentes na Web, *e-mails*, ou aplicações de *mensagens instantâneas* (IM). Além disso, as aplicações de negócios podem ser programadas para agirem de acordo com as descobertas desses sistemas de BI em tempo real. Por exemplo, uma aplicação de *gestão da cadeia de fornecimento* (SCM) pode automaticamente fazer um pedido de mais “produtos X” quando o estoque em tempo real ficar abaixo de certo limite. Ou uma aplicação de *gerenciamento de relacionamento com o cliente* (CRM) pode solicitar que um representante do

serviço ao cliente e um funcionário do controle de crédito verifiquem um cliente que tenha feito um pedido *online* superior a \$ 10.000.

Uma abordagem ao BI em tempo real usa o modelo de DW dos sistemas tradicionais. Nesse caso, os produtos de fornecedores de plataformas inovadoras de BI (como Ascential ou Informatica) oferecem uma solução orientada ao serviço e quase em tempo real, que preenche o DW com muito mais rapidez que uma típica atualização noturna por *extração, transformação e carga* (ETL; Capítulo 2). Uma segunda abordagem, comumente chamada *monitoramento de atividades de negócios* (BAM; Capítulo 5) é adotada por fornecedores de BAM exclusivamente e/ou de híbridos BAM-*middleware* (como Savvion, Iteration Software, Vitria, webMethods, Quantive, Tibco ou Vineyard Software). Ela se desvia totalmente do DW e usa os *Web services* ou outras formas de monitoramento para descobrir eventos essenciais de negócios. O uso da Web facilita o BI em tempo real (Giovino: 2002). Esses monitores de *software* (ou agentes inteligentes) podem ser colocados em um servidor separado da rede ou nos próprios bancos de dados de aplicações transacionais, e podem usar abordagens baseadas em eventos ou em processos para medir e monitorar os processos operacionais de forma proativa e inteligente. Para mais detalhes, consulte Thompson e Jakovljeric (2005).

Desenvolvendo ou adquirindo sistemas de BI

Atualmente, muitos fornecedores oferecem ferramentas diversificadas, algumas das quais são totalmente pré-programadas (chamadas *shells*): basta apenas inserir os números. É possível comprá-las ou arrendá-las. Para ver uma lista de produtos em versão *demo*, visite dmreview.com/resources/demos.cfm. Na verdade, quase todos os sistemas de BI são construídos com *shells*, frequentemente de fornecedores terceirizados. O problema que as empresas enfrentam é qual alternativa selecionar: compra, arrendamento* ou construção. Cada uma dessas tem diversas opções. Um dos principais critérios a serem considerados é a justificativa e a análise de custo/benefício.

Justificativa e análise de custo/benefício

Com o aumento do número de aplicações de BI em potencial vem a necessidade de justificá-las e priorizá-las. Essa não é uma tarefa fácil devido ao grande número de benefícios intangíveis. No entanto, a partir de 2003, as diretorias de empresas começaram a pressionar para que se fizesse uma análise de custo/benefício em qualquer sistema de preço superior a certo nível. Há várias maneiras de fazê-lo.

Segurança e proteção da privacidade

A segurança é um aspecto extremamente importante do desenvolvimento de qualquer sistema computadorizado contendo dados que possam ter valor estratégico, principalmente de BI. A privacidade de funcionários e clientes também precisa ser protegida.

Integração entre sistemas e aplicações

Exceto por algumas aplicações pequenas, todas as aplicações de BI devem ser integradas a outros sistemas, como bancos de dados, sistemas legados, sistemas empresariais (principalmente ERP e CRM), comércio eletrônico (lado vendedor, lado comprador) e muitos outros. Além disso, as aplicações de BI normalmente são conectadas à Internet e, muitas vezes, também aos sistemas de informações de parceiros de negócio. Ademais, as ferramentas de BI, às vezes, precisam ser integradas umas às outras, criando sinergia.

A necessidade de integração forçou os fornecedores de *software* a adicionarem recursos a seus produtos continuamente. Os clientes que compram um pacote de *software* completo nego-

* N. de R.: Em inglês, *leasing*.

ciam com apenas um fornecedor e não têm de lidar com conectividade do sistema. Mas eles podem perder a vantagem de criar sistemas compostos pelos melhores componentes do mercado.

Questões de revisão da Seção 1.7

1. Descreva os principais tipos de usuários de BI.
2. Liste alguns dos tópicos da implementação abordados pelo relatório do Gartner.
3. Liste alguns outros fatores de sucesso do BI.
4. Descreva as questões envolvidas na aquisição ou construção de uma aplicação de BI.
5. Por que é difícil justificar as aplicações de BI?
6. Que tipos de integração são necessárias na implementação do BI?

1.8

CONCLUSÃO: O *BUSINESS INTELLIGENCE* HOJE E AMANHÃ

No ambiente de negócios altamente competitivo de hoje, a qualidade e pontualidade da informação de negócios para uma organização não é a escolha entre lucro ou perda: pode ser uma questão de sobrevivência ou falência. Nenhuma organização pode negar os inevitáveis benefícios do BI. Relatórios recentes de analistas do setor mostraram que, nos próximos anos, milhões de pessoas usarão ferramentas visuais de BI e análise de BI todos os dias (Baum: 2006).

As organizações atuais estão obtendo mais valor do BI por estenderem as informações a muitos tipos de funcionários, e assim maximizarem o uso dos ativos de dados existentes.

Ferramentas de visualização que incluem *dashboards* são usadas por produtores, varejistas, governos e órgãos especiais. Mais e mais ferramentas de análise específicas do setor irão invadir o mercado, para executar praticamente qualquer tipo de análise e facilitar a tomada de decisões informada, desde o nível mais estratégico até o nível dos usuários.

Uma tendência em potencial envolvendo a BI é sua possível *fusão com a inteligência artificial* (AI). A AI tem sido utilizada em aplicações de negócios desde a década de 1980, e é amplamente usada na resolução de problemas complexos e em técnicas de suporte à decisão de aplicações de negócios em tempo real. Não levará muito tempo até a fusão entre aplicações de AI e BI, e isso dará início a uma nova era nos negócios. Os ADSs são o primeiro passo nessa direção. Para permitir a integração, os fornecedores de BI estão começando a usar a arquitetura orientada a serviços (Lal: 2005) e a integração de informações corporativas (EII). (Consulte Thompson and Jakovljevic: 2005.)

O *business intelligence* está abrindo suas asas para abraçar todos, desde pequenas e médias empresas até grandes organizações. Assim como há grandes protagonistas de BI para grandes corporações, há também os protagonistas dos pequenos nichos que servem as empresas médias e pequenas. Ferramentas de análise para funções muito específicas também estão entrando no mercado, e elas ajudarão algumas empresas a adotar apenas a análise em vez da implementação completa de BI baseada em *data warehouse*.

O BI aproveita os componentes já desenvolvidos e instalados das tecnologias de TI para ajudar as empresas a fazerem bom uso de seus investimentos em TI, bem como a utilizarem os valiosos dados armazenados em sistemas legados e transacionais. Para muitas empresas de grande porte que já gastaram milhões de dólares construindo *data warehouses* e *data marts* (como para uso em CRM), agora é a hora certa de construir BI como a próxima etapa para garantir o benefício completo de seu investimento, que terá impacto diretamente no ROI. No entanto, embora alguns componentes da BI, como o DW, possam mudar (p. ex., os dados podem ser armazenados online), a necessidade de realizar BI no ambiente de negócios em rápida modificação aumentará, fazendo do BI uma necessidade. Para saber mais sobre o futuro do BI, veja Lal (2005).

RECURSOS, LINKS E CONEXÃO COM A TERADATA UNIVERSITY NETWORK

O uso deste e da maioria dos outros capítulos do livro pode ser aprimorado pelos itens a seguir.

Recursos e links

Os principais recursos e *links* que recomendamos são:

- Bez Systems, **bez.com**
- IT Toolbox Business Intelligence Knowledgebase, **businessintelligence.ittoolbox.com**
- Datawarehousing.com, **datawarehousing.com**
- Business Intelligence and Data Warehousing Insight Portal, **datawarehousingonline.com**
- Decision Support Systems Resources, **dssresources.com**
- TD Magazine, **teradatamagazine.com**
- The *Data Warehouse* Institute, **dw-institute.com** ou **tdwi.org**
- The DMReview, **dmreview.com**
- OLAP Report, **olapreport.com**
- B-Eye-Network, **b-eye-network.com**
- IS World, **isworld.org**

Casos

Todos os principais fornecedores de BI têm histórias interessantes de sucesso com os clientes (p. ex., MicroStrategy, Microsoft, Oracle, IBM, Hyperion, Cognos, Exsys, SAS, Fairisaac, Business Objects, SAP e Information Builders). Casos acadêmicos estão disponíveis em Harvard Business School Case Collection, **bpir.com** (Massey University), **idea-group.com**, Ivy League Publishing (**ivyip.com**), **icmr.icfai.org/casestudies/icmr_case_studies.htm**, **knowledgestorm.com**, etc.

Fornecedores, produtos e demos

Há uma lista abrangente de versões *demo* em **dmreview.com/resources/demos.cfm**. Informações sobre produtos, arquitetura e *software* podem ser encontradas em **dsslab.com**.

Revistas e periódicos

CIO Insights (**ciainsight.com**)

Technologyevaluation.com (uma revista *online*)

Baseline Magazine (**baselinemag.com**)

Business Intelligence Journal (de **tdwi.org**)

Advisor.com (um jornal *online*)

Conexão com a Teradata University Network

Nosso livro está intimamente ligado aos recursos gratuitos oferecidos pela Teradata University Network (TUN; consulte **teradatauniversitynetwork.com** e o Arquivo *online* W1.5 para ler a descrição). O portal da TUN é dividido em duas partes principais: uma para estudantes – **teradatastudentnetwork.com** –, a outra para o corpo docente. Estamos conectados a este portal através dos exercícios ao final de cada capítulo. Proporcionamos exercícios interativos usando *software* e outros materiais (como casos, seminários na Webinars) disponíveis na TUN.

Material online

No site da Bookman Editora, www.bookman.com.br, há material complementar para os estudantes e para professores, além de um capítulo *online*.

Destaques do capítulo

- O ambiente de negócios está se tornando complexo e mudando rapidamente, o que deixa as decisões mais difíceis.
- As empresas devem reagir e se adaptar prontamente aos ambientes em mudança tomando decisões melhores e mais velozes.
- O intervalo para tomar decisões está diminuindo, enquanto sua natureza global está se expandindo: isso exige desenvolvimento e uso de sistemas de suporte computadorizado às decisões.
- A taxa de informatização aumenta rapidamente, assim como seu uso no suporte ao BI.
- A tomada de decisões gerenciais se tornou complexa. Métodos intuitivos e de tentativa e erro podem não ser suficientes.
- Os sistemas de suporte à gerência são tecnologias criadas para dar apoio ao trabalho gerencial. Podem ser usados individualmente ou em conjunto.
- O suporte computadorizado para os gerentes é, com frequência, essencial à sobrevivência das organizações.
- O suporte automatizado às decisões é oferecido hoje em muitos setores em casos de essas serem repetitivas, baseadas em regras de negócios (como a precificação de itens).
- Os métodos de *business intelligence* utilizam um repositório central (*data warehouse*) que permite eficiência em *data mining*, processamento analítico *online*, BPM e visualização de dados.
- A arquitetura do *business intelligence* inclui o *data warehouse*, ferramentas de análise de negócios usadas pelos usuários finais e uma interface de usuário (como um *dashboard*).
- Sistemas de *business* (ou *corporate*) *performance management* são usados por muitas organizações para monitorar o desempenho, compará-lo a padrões e metas e exibi-lo graficamente (p. ex., por *dashboards*) aos gerentes, executivos e analistas.
- *Data mining* é uma ferramenta para descobrir informações e relações em uma grande quantidade de dados.

Termos-chave

- | | | |
|--|--|---|
| • <i>business intelligence</i> (BI) | • <i>data mining</i> | • processamento de transações |
| • <i>business performance management</i> (BPM) | • governança do BI | • <i>online</i> (OLTP) |
| • <i>dashboard</i> | • processamento analítico <i>online</i> (OLAP) | • sistemas automatizados de decisão (ADS) |

QUESTÕES PARA DISCUSSÃO

1. Explique os cinco Ws do processo geral de criação e uso da inteligência (do inglês *who, what, when, where* e *why*: quem, o quê, quando, onde e por quê).
2. Por que não é suficiente avaliar os benefícios tangíveis de ROI ou TCO para um investimento em DW? Por que é difícil?
3. O que é governança do BI?
4. Diferencie coleta de inteligência e espionagem.
5. Discuta a relação entre sucesso de BI e BI em tempo real e sob demanda.
6. Relacione a coleta de inteligência à inteligência competitiva.
7. Explique os papéis dos diferentes usuários de BI.
8. Quais são as questões da alta administração que devem ser avaliadas antes do lançamento de uma iniciativa de DW/BI?
9. Quais são as funções mais importantes de um centro de competência de *business intelligence* (BICC) em uma organização?
10. Por que é necessário usar a metodologia de BPM?
11. Discuta as questões mais importantes da implementação de BI.

Exercícios

Exercícios da Teradata Student Network (TSN) e outros exercícios interativos

1. Entre em teradatastudentnetwork.com (seu instrutor emitirá as informações de registro). Obtenha uma senha com seu instrutor e estude o conteúdo do *site*. Monte uma lista de todo o material disponível. Você receberá tarefas relacionadas ao *site*. Prepare uma lista de 20 itens do *site* que poderiam ser vantajosos para você.
2. Entre em teradatastudentnetwork.com (TSN) e encontre o documento intitulado “Data Warehousing Supports Corporate Strategy at First American Corporation” (de Warson, Wixom e Goodhue). Leia-o e responda as seguintes questões:
 - a. Que elementos impulsionaram o projeto de DW/BI na empresa?
 - b. Quais vantagens estratégicas foram obtidas?
 - c. Quais vantagens operacionais e táticas foram obtidas?
 - d. Quais foram os FCS da implementação?
3. Entre na TSN e encontre o seminário na Webinar chamado “Enterprise Business Intelligence: Strategies and Technologies for Deploying BI on Large Scale” (de Eckerson e Howson). Visualize o seminário na Webinar e responda as seguintes questões:
 - a. Quais são as vantagens de implantar BI para o grande público?
 - b. Quem são os usuários de BI em potencial? O que cada tipo de usuário tenta obter?
 - c. Que lições de implementação você aprendeu do seminário?
4. Entre na TSN e encontre o caso Harrah’s (de Watson e Volonino). Responda:
 - a. Quais eram os objetivos do projeto?
 - b. Qual foi o papel do DW?
 - c. Que tipos de análises foram usados?
 - d. Que vantagens estratégicas o BI oferece?
 - e. Qual é a importância de um executivo de inovação?

Para mais informações sobre a Harrah’s, consulte G. Loveman (May 2003), “Diamonds in the Data”, *Harvard Business Review*, e G. Evans (June 2006), “The Big Payoff”, *Teradata Magazine*.

Tarefas em grupo e interpretação de papéis

1. Encontre informações sobre o uso proativo de computadores para dar suporte ao processamento analí-

tico *online* em comparação aos sistemas de processamento de transações (SPT). Cada integrante do grupo deve escolher uma aplicação em um setor diferente (varejo, bancos, seguros, alimentação, etc.). Não deixe de incluir os impactos da Web/Internet. Resuma as descobertas e aponte as semelhanças e diferenças das aplicações. Como fontes, use empresas nas quais os estudantes trabalham, revistas especializadas, *newsgroups* da Internet e *websites* de fornecedores. Por fim, prepare uma apresentação sobre as descobertas.

2. Visite fairisaac.com, ilog.com e pega.com. Veja as demonstrações. Prepare uma lista de suportes automatizados à decisão por setor e por área funcional. Especifique que tipos de decisões são automatizados.

Exercícios na Internet

1. Acesse fairisaac.com e identifique cinco problemas em diferentes setores e cinco problemas em diferentes áreas funcionais que podem receber suporte de ADSs.
2. Acesse sap.com e oracle.com e encontre informações sobre como os *software* de planejamento de recursos empresariais (ERP) auxiliam os tomadores de decisão. Relacione ao BI. Escreva um relatório.
3. Acesse intelligententerprise.com. Para cada termo-chave citado neste capítulo, encontre algum evento importante noticiado no *site* e prepare um relatório.
4. Entre em cognos.com e businessobjects.com e compare os recursos de seus produtos de BI em um relatório.
5. Acesse microsoft.com. Examine suas ofertas de BI.
6. Acesse oracle.com. Verifique suas ofertas de BI. Como elas se relacionam ao *software* de ERP da Oracle?
7. Entre em microstrategy.com e encontre informações sobre os cinco estilos de BI. Organize uma tabela resumida para cada estilo.
8. Visite hyperion.com e descubra quais são seus principais produtos. Relacione-os às tecnologias de suporte mencionadas neste capítulo.
9. Entre em businessobjects.com e localize as necessidades de negócios da BI e as soluções oferecidas para as necessidades. Prepare um relatório.

CASO DE APLICAÇÃO DE FINAL DO CAPÍTULO

BNSF usa BI para melhorar serviço, agilidade, eficiência e lucratividade

Sediada em Fort Worth, Texas, a BNSF foi criada a partir da fusão das ferrovias Burlington Northern e Santa Fe Pacific em 1995. A BNSF opera uma das maiores redes ferroviárias da América do Norte: emprega 38.000 pessoas em 52.300 quilômetros de linhas no Estados Unidos e no Canadá. As locomotivas Dash 9 da empresa usam um sistema de controle com microprocessador integrado que oferece diagnósticos embutidos, bem como outros sistemas que melhoram a economia de combustível, reduzem as emissões e aumentam a vida útil do motor. A BNSF é uma das maiores empresas de transporte do mundo com tráfego intermodal e tem a maior ferrovia para carregamento de grãos, além de transportar carvão o suficiente para gerar mais de 10% da eletricidade produzida nos Estados Unidos.

O PROBLEMA

Os clientes da BNSF exigem uma abordagem integrada à informação ao longo da cadeia de fornecimento, para que não tenham de negociar separadamente com consolidadores, corretores, barcos a vapor e transportadoras. Com aproximadamente 100.000 carros sempre ativos em rotas na metade oeste dos Estados Unidos, uma mudança ambiental como inundações ou nevascas em uma estrada principal provoca alterações na rota e exige notificações a todos os clientes o mais rápido possível, a fim de manter o cumprimento dos prazos. Para isso, a BNSF precisava entender melhor seus clientes, utilizar seus ativos de forma eficiente, aumentar a renda e reduzir os custos.

A previsão e os orçamentos, atividades importantes relativas à administração, eram feitas usando uma quantidade excessiva de planilhas nas quais as informações eram inseridas à mão. O resultado era um processo tedioso e demorado. A empresa depositava muita confiança em métricas no processo de tomada de decisões, e precisava de um método poderoso para usar essas métricas em diversos relatórios e de recursos que simplificassem o processo.

A empresa também precisava oferecer uma fonte única de informações referentes aos carregamentos, desde o momento do pedido até o recebimento do pagamento, para reduzir os custos e lidar com o processo de faturamento de forma mais eficaz.

A SOLUÇÃO

Os princípios de negócio a que a BNSF obedece, como consequência do uso expandido do BI, incluem: 1) oferecer uma versão única da verdade, 2) armazenar os detalhes, 3) proporcionar maior valor através da produtividade, 4) permitir melhor análise. Conforme se desenvolvem aplicações, esses princípios são guias para garantir que a empresa forneça o máximo valor agregado.

Ferramentas, processos e metodologias que permitem facilmente aos clientes fazerem interface com a BNSF determinaram o desenvolvimento de **bnsf.com**, uma suíte de ferramentas baseadas na Web usadas pelos clientes e integradas a seus sistemas.

Em 2002, a BNSF projetou o OnTrack, um sistema de previsão e planejamento que usa o *Teradata Warehouse* para armazenar os dados, tecnologia de *front-end* construída a partir do Essbase da Hyperion, e algumas telas personalizadas desenvolvidas com Java. A ferramenta de relatórios da Web IBM DB2 Alphablox e o suplemento do Excel da Hyperion oferecem recursos de geração de relatórios. O OnTrack permite rápidas situações “e se”, e usa regras e definições de negócios para acelerar o processo e oferecer flexibilidade (pelo modo “e se”).

O Corporate Dashboard fornece um local único para que executivos e analistas encontrem informações coordenadas e integradas e trabalhem em um painel comum. A BNSF construiu os sistemas internamente usando dados extraídos dos sistemas *Teradata Warehouse*, DB2 e SAS, com o Essbase para a parte de OLAP e o Alphablox para distribuição na Web. A ferramenta tem três visualizações principais dentro do Corporate Dashboard. Por meio da primeira, que usa um código de cores de “sinal vermelho” para apontar rapidamente áreas de exceção, um gerente geral pode chegar rapidamente aos detalhes das áreas de exceção para delinear o local exato no qual ocorre o problema. Mais detalhes ou correlações com outros pontos de dados estão disponíveis na visualização em que o usuário consegue rapidamente localizar os problemas, detalhar essas áreas e obter várias visualizações diferentes relacionadas a outros tipos de métricas. Para uma análise ainda mais profunda, o usuário pode retalhar os elementos de dados do *data*

warehouse e de outras fontes para obter uma perspectiva realmente abrangente.

O sistema Revenue Single Source aprimora a capacidade de alterar conhecimentos de embarque, elimina a redundância e ajuda a processar os pagamentos com mais rapidez e precisão. Ele está disponível por meio de um sistema da Web que usa mecanismos baseados em regras, busca padrões para processamento e combina alertas de tolerância e notificação de eventos enquanto utiliza um *data warehouse* ativo. A adição de raciocínio baseado em casos analisa a qualidade de uma decisão.

OS RESULTADOS

O sistema OnTrack oferece maior flexibilidade, permitindo que os gerentes de orçamentos criem suas previsões e também planejem e façam previsões no modo “e se”. Ao mudarem uma variável, os usuários criam uma previsão alternativa que pode ser usada para comparações e contrastes, para que se chegue à melhor previsão. O novo sistema é mais automatizado e padronizado, o que facilita e agiliza seu uso. Funções padrões e *ad hoc* de relatórios foram ampliadas para 80 usuários avançados do OnTrack e os relatórios são distribuídos para 300 usuários.

O Corporate Dashboard permite um tempo menor de reação. A ferramenta estabelece uma base para a análise preditiva e oferece uma percepção dos negócios concentrando-se em indicadores-chave, apontando rapidamente as áreas problemáticas, simplificando

a navegação por essas áreas, e resumindo os fatores de dados críticos que agilizam o tempo de análise e permitem mais tempo para avaliar e tomar decisões.

Em 2004, a BNSF foi nomeada Premier Technology Leader (principal líder tecnológico) pela revista *Computerworld*, e recebeu prêmios da Target e da Toyota Logistic Services pelo serviço ao cliente e do Wal-Mart de Carrier of the Year (transportador do ano). Em 2006, a BNSF ganhou o prêmio relativo à gestão do desempenho financeiro da Ventana Research, a principal empresa de pesquisa em gestão do desempenho e serviços de consultoria.

Fontes: compilado de K. Schwartz, “Seeing Around the Bend at BNSF”, *Teradata Magazine Online*, March 2005, teradata.com/t/page/131972/index.html (acessado em agosto de 2006); NCR.COM. “Teradata Customer BNSF Railway Recognized for Leadership in Finance and Performance Management”, August 1, 2006, ncr.com/en/media_information/2006/aug/pr080106.htm (acessado em agosto de 2006).

QUESTÕES SOBRE O CASO

1. Descreva os elementos que motivaram a BI neste caso. Liste os problemas e necessidades específicos.
2. Como as soluções de BI ajudaram a resolver esses problemas?
3. Relacione a solução à análise de negócios.
4. Descreva a função do *dashboard*.

Referências

- Baum, D. (March/April 2006). “The Face of Intelligence.” *ORACLE Magazine*.
- Biere, M. (2003). *Business Intelligence for the Enterprise*. Upper Saddle River, NJ: Prentice Hall.
- Business Objects (2002). “France Telecom BI in Action.” *Business Objects’ Case Study*. businessobjects.com/pdf/success/France_Telecom.pdf (accessed February 2007).
- Coveney, M., et al. (2003). *The Strategy Gap*. Hoboken, NJ: John Wiley and Sons.
- Davenport, T. (January 2006). “Competing on Analytics.” *Harvard Business Review*.
- Davenport, T. H., and J.G. Harris. (2005, Summer). “Automated Decision Making Comes of Age.” *MIT Sloan Management Review*.
- Dignan, K. G. (September 2003). “Project Portfolio Management for BI.” *TDWI*.
- Eckerson, W. (2003). *Smart Companies in the 21st Century: The Secrets of Creating Successful Business Intelligent Solutions*. Seattle, WA: The Data Warehousing Institute.
- Evans, G. (June 2006). “The Big Payoff.” *Teradata Magazine*.
- Gartner Inc. (2004). *Using Business Intelligence to Gain a Competitive Edge*. A special report. Gartner, Inc.: Stamford CT.
- Giovinazzo, W. A., (2002). *Internet-Enabled Business Intelligence*. Upper Saddle River, NJ: Prentice Hall.
- Heinrich, C. (2005). *RFID and Beyond*. Indianapolis IN: Wiley Publishing Inc.
- Huber, G. (2003). *The Necessary Nature of Future Firms: Attributes of Survivors in a Changing World*. San Francisco: Sage Publication.
- Imhoff, C., and R. Pettit (2004). “The Critical Shift to Flexible Business Intelligence.” White paper, Intelligent Solutions, Inc.
- Inmon, W. H. (2005). *Building Data Warehouses*, 4th ed. New York: John Wiley.
- Kelly, C. (June 14, 2001). “Calculating Data Warehousing ROI.” *Data Warehousing and Business Intelligence*.

- Krizan, L. (June 1999). *Intelligence Essentials for Everyone*. Washington DC: Joint Military Intelligence College (occasional paper number six), Department of Defense.
- Lal, V. (2005). *The Future of Business Intelligence*. White paper. Santa Clara, CA: Hyperion Solutions Corp.
- Loschin, D. (2003). *Business Intelligence: The Savvy Manager's Guide*. San Francisco, CA: Morgan Kaufmann.
- Matney, D., and D. Larson (Summer 2004). "The Four Components of BI Governance." *Business Intelligence Journal*.
- Raisinghani, M. (2004). *Business Intelligence in the Digital Economy*. Hershey PA: The Idea Group.
- Rogalski, S., and F. Lin (November 2003). "Business Intelligence: The Impact of SOX on Financial Reporting." *DM Review Magazine*.
- Thompson, O. (October 2004). "Business Intelligence Success, Lessons Learned." *TechnologyEvaluation.com*.
- Thompson, O., and P. J. Jakovljevic (July 16, 2005). "Business Intelligence Status Report" (a seven-part series). *TechnologyEvaluation.com*. technologyevaluation.com/research/center/BusinessIntelligence/index.asp.
- Williams, S., and N. Williams. (Fall 2003). "The Business Value of Business Intelligence." *Business Intelligence Journal*.
- Zaman, M. (January 2005). "Business Intelligence: Its Ins and Outs." technologyevaluationcenters.com



Data Warehousing

Objetivos de aprendizado

- ◆ Compreender as definições e os conceitos básicos dos *data warehouses*.
- ◆ Compreender as arquiteturas de *data warehousing*.
- ◆ Descrever os processos usados no desenvolvimento e gerenciamento dos *data warehouses*.
- ◆ Explicar as operações de *data warehousing*.
- ◆ Explicar o papel dos *data warehouses* no suporte à decisão.
- ◆ Explicar a integração de dados e os processos de extração, transformação e carga (ETL).
- ◆ Descrever o *data warehousing* em tempo real (ativo).
- ◆ Compreender as questões de administração e segurança do *data warehouse*.

O conceito de *data warehousing* existe desde o final dos anos 1980. Este capítulo oferece a base para um importante tipo de banco de dados, chamado de *data warehouse*, que é cada vez mais usado no suporte à decisão e que proporciona aprimorados recursos de análise. Discutiremos o *data warehousing* nas seguintes seções:

- 2.1 Vinheta de abertura: a Continental Airlines voa alto com seu *data warehouse* em tempo real
- 2.2 Definições e conceitos de *data warehousing*
- 2.3 Visão geral do processo de *data warehousing*
- 2.4 Arquiteturas do *data warehousing*
- 2.5 Integração de dados e processos de extração, transformação e carga (ETL)
- 2.6 Desenvolvimento de *data warehouses*
- 2.7 *Data warehousing* em tempo real
- 2.8 Questões de administração e segurança do *data warehouse*
- 2.9 Recursos, *links* e conexão com a Teradata University Network

2.1

VINHETA DE ABERTURA: A CONTINENTAL AIRLINES VOA ALTO COM SEU DATA WAREHOUSE EM TEMPO REAL

À medida que o *business intelligence* (BI) torna-se um componente crítico das operações diárias, são cada vez mais implementados *data warehouses* em tempo real, que oferecem aos usuários finais

atualizações rápidas e alertas gerados a partir dos sistemas transacionais. O *data warehousing* e o BI em tempo real, que apoiam seu dinâmico plano de negócio Go Forward, ajudaram a Continental Airlines a alterar seu status no setor de “pior para primeiro” e depois de “primeiro para favorito”. A Continental Airlines é uma líder em BI em tempo real. Em 2004, a empresa ganhou o prêmio Best Practices and Leadership (melhores práticas e liderança) do Data Warehousing Institute.

Grandes problemas

A Continental Airlines foi fundada em 1934, com um monomotor Lockheed, no sudoeste dos EUA. Desde 2006, a Continental (Houston) é a quinta maior companhia aérea dos Estados Unidos e a sétima maior do mundo. Ela tem a mais ampla rede global de rotas entre todas as empresas de aviação norte-americanas, com mais de 2.300 partidas diárias para mais de 227 destinos.

Em 1994, a Continental enfrentava sérios problemas financeiros. Por duas vezes ela havia solicitado proteção contra o Capítulo 11 de falência da legislação americana e estava rumo à terceira falência, provavelmente a final. As vendas de passagens estavam prejudicadas porque o desempenho dos fatores importantes para os clientes era lastimável, incluindo uma porcentagem baixa de partidas pontuais, problemas freqüentes na chegada das bagagens e muitos clientes mandados embora devido ao *overbooking*.

Solução

O renascimento da Continental teve início em 1994, quando Gordon Bethune assumiu a presidência e iniciou o plano Go Forward, que consistiu na implementação simultânea de quatro partes inter-relacionadas. Bethune teve como alvo a necessidade de melhorar os indicadores de desempenho valorizados pelos clientes através de um melhor entendimento das necessidades dos clientes e de suas percepções sobre o valor dos serviços que eram e poderiam ser oferecidos. As práticas de gestão financeira também foram alvo de uma reforma considerável. Já em 1998, a companhia possuía bancos de dados separados para *marketing* e operações, todos hospedados e gerenciados por fornecedores externos. O processamento de consultas e o fomento de programas de *marketing* para seus maiores clientes eram demorados e ineficientes. Além disso, as informações de que os funcionários necessitavam para tomar decisões rápidas simplesmente não estavam disponíveis. Em 1999, a Continental decidiu integrar suas fontes de dados operacionais, de *marketing*, de TI e de receita em um único *data warehouse* empresarial (EDW) interno. O *data warehouse* ofereceu diversos benefícios valiosos e antecipados.

Assim que a Continental voltou a ter lucros e ficou em primeiro lugar no setor aéreo em muitas métricas de desempenho, Bethune e sua equipe administrativa elevaram o nível expandindo a visão. Em vez de simplesmente ter um desempenho melhor, a intenção era que a Continental fosse a companhia aérea favorita dos clientes. O plano Go Forward estabeleceu formas mais práticas de uma empresa tornar-se a preferida entre os viajantes. A tecnologia passou a ser cada vez mais crítica no suporte a essas iniciativas. No princípio, ter acesso às informações históricas e integradas bastava. Isso rendia um bom valor estratégico. Porém, foi mais e mais necessário que o *data warehouse* oferecesse informações acionáveis em tempo real para dar suporte à tomada de decisões táticas e aos processos comerciais em toda a corporação.

Por sorte, a equipe de *warehouse* já esperava por isso e fez os arranjos para a mudança em tempo real. Desde o início, a equipe havia criado uma arquitetura para lidar com alimentação de dados ao *warehouse* em tempo real, extrações de dados dos sistemas legados e consultas táticas que exigiam tempos de resposta quase imediatos. Em 2001, foram disponibilizados dados em tempo real a partir do *warehouse*; a quantidade armazenada cresceu rapidamente. A Continental move dados em tempo real (que variam de dados do minuto até dados da hora) sobre clientes, reservas,

check-ins, operações e vôos, de seus principais sistemas operacionais ao warehouse. As aplicações em tempo real da Continental incluem:

- Gestão e contabilidade da receita
- Gerenciamento de relacionamento com o cliente (CRM)
- Operações e folha de pagamento da equipe
- Segurança e fraude
- Operações de vôo

Benefícios

Somente no primeiro ano, após a implementação do projeto de *data warehouse*, a Continental identificou e eliminou mais de US\$ 7 milhões em fraude e reduziu os custos em US\$ 41 milhões. Com um investimento de US\$ 30 milhões em *hardware* e *software* durante seis anos, a Continental chegou a mais de US\$ 500 milhões em aumento da receita e economia em *marketing*, detecção de fraudes, previsão de demanda e acompanhamento, além de melhorar a gestão do centro de dados. A visão única, integrada e confiável da empresa (i.e. a versão única da verdade) levou a uma tomada de decisões melhor e mais rápida.

A Continental é agora identificada como uma líder de BI em tempo real, com base em sua arquitetura escalonável e extensível, decisões práticas sobre quais dados são captados em tempo real, fortes relações com usuários finais, equipe de *data warehouse* pequena e altamente competente, ponderação sensata sobre as exigências do suporte à decisão estratégica e tática, compreensão das sinergias entre o suporte à decisão e as operações e alteração de processos de negócios para usarem dados em tempo real. (Para ver uma amostra de tela do sistema da Continental, visite teradata.com/t/page/139245/.)

Fontes: adaptado de H.Wixom, J. Hoffer, R. Anderson-Lehman, and A. Reynolds, “Real-Time Business Intelligence: Best Practices at Continental Airlines”, *Information Systems Management Journal*, Vol. 23, No. 1, Winter 2006, pp. 7–18; R. Anderson-Lehman, H.Watson, B.Wixom, e J. Hoffer, “Continental Airlines Flies High with Real-Time Business Intelligence”, *MIS Quarterly Executive*, Vol. 3, No. 4, December 2004, pp. 163–176 (disponível em teradatauniversitynetwork.com); H.Watson, “Real Time:The Next Generation of Decision-Support Data Management”, *Business Intelligence Journal*, Vol. 10, No. 3, 2005, pp. 4–6; M. Edwards, “2003 Best Practices Awards Winners: Innovators in Business Intelligence and Data Warehousing”, *Business Intelligence Journal*, Vol. 8, No. 4, Fall 2003, pp. 57–64; R.Westervelt, *Continental Airlines Builds Real-Time Data Warehouse*, August 20, 2003, searchoracle.techtarget.com; R. Clayton, “Enterprise Business Performance Management: Business Intelligence + Data Warehouse = Optimal Business Performance”, *Teradata Magazine*, September 2005, teradata.com/t/page/139245/; and The Data Warehousing Institute, *2003 Best Practices Summaries: Enterprise Data Warehouse*, 2003, tdwi.org/display.aspx?ID=6749.

Questões sobre a vinheta de abertura

1. Descreva os benefícios da implementação da estratégia Go Forward da Continental.
2. Explique por que é importante que uma companhia aérea use um *data warehouse* em tempo real.
3. Examine a amostra de tela do sistema no endereço teradata.com/t/page/139245/. Descreva como ela pode auxiliar o usuário a identificar problemas e oportunidades.
4. Identifique as principais diferenças entre o *data warehouse* tradicional e um *data warehouse* em tempo real, como o que foi implementado na Continental.
5. Que vantagem estratégica a Continental pode obter do sistema em tempo real em oposição a um sistema de informação tradicional?

O que podemos aprender com esta vinheta?

A vinheta de abertura ilustra o valor estratégico da implementação de um *data warehouse*, juntamente com seus métodos de suporte de BI. De uma das piores companhias aéreas, beirando a falência, a Continental conseguiu se tornar uma empresa de aviação de primeira classe e em

pouco tempo, gerando aumento significativo de receita e redução de custos. A redução de custos no primeiro ano da implementação mais do que pagou o investimento de seis anos em sistemas. O *data warehouse* integrou vários bancos de dados de toda a organização em uma única unidade interna na empresa, o que gerou uma versão única da verdade da companhia, colocando todos os funcionários no mesmo patamar. Além disso, os dados foram disponibilizados em tempo real para os tomadores de decisão: isso acabou criando uma vantagem competitiva estratégica para a continental no setor. A principal lição aqui é que um *data warehouse* em tempo real criado para a empresa, combinado a uma estratégia para seu uso no suporte à decisão, pode melhorar o uso dos dados e oferecer benefícios financeiros gigantescos para uma organização.

2.2

DEFINIÇÕES E CONCEITOS DE DATA WAREHOUSING

Usar o *data warehousing* em tempo real juntamente com sistemas de suporte à decisão (DSS) e ferramentas de BI é uma forma relevante de conduzir os processos de negócios. A vinheta de abertura demonstra uma situação em que um *data warehouse* em tempo real deu suporte à tomada de decisões, por intermédio da análise de grandes quantidades de dados de várias fontes, para propiciar resultados rápidos e apoiar os processos críticos. Com fluxos de dados em tempo real, a Continental pode ver o estado atual de seus negócios e identificar seus problemas, o que é o primeiro passo rumo à resolução analítica desses. Além disso, os clientes podem obter o status em tempo real de vôos e outras informações descritivas; desta forma, o sistema oferece também uma vantagem competitiva expressiva sobre os concorrentes.

Os tomadores de decisão necessitam informações concisas e confiáveis sobre operações atuais, tendências e mudanças. Os dados são constantemente fragmentados com o uso de sistemas operacionais diferentes, e assim os gerentes tomam decisões com informações parciais, na melhor das hipóteses. O *data warehouse* supera esse obstáculo acessando, integrando e organizando os principais dados operacionais de uma forma consistente, confiável, pontual e prontamente disponível onde for necessário.

O que é um *data warehouse*?

Em poucas palavras, um *data warehouse* (DW) é um conjunto de dados produzido para oferecer suporte à tomada de decisões; é também um repositório de dados atuais e históricos de possível interesse aos gerentes de toda a organização. Os dados normalmente são estruturados de modo a estarem disponíveis em um formato pronto para as atividades de processamento analítico (p. ex. processamento analítico *online* [OLAP], *data mining*, consultas, geração de relatórios, outras aplicações de suporte à decisão). Portanto, um *data warehouse* é uma coleção de dados orientada por assunto, integrada, variável no tempo e não-volátil, que proporciona suporte ao processo de tomada de decisões da gerência.

Características do *data warehousing*

Uma forma comum de apresentar o *data warehousing* é recorrer às suas características fundamentais (consulte Inmon: 2005):

- *Orientado por assunto.* Os dados são organizados por assunto pormenorizado, como vendas, produtos ou clientes, e contêm apenas as informações relevantes ao suporte à decisão. A orientação por assunto permite que os usuários determinem não só como está o desempenho de sua empresa, mas também permite determinar o porquê deste desempenho. Um *data warehouse* difere de um banco de dados operacional no sentido de que estes, em sua maioria, são orientados por produto e ajustados para lidar com transações que atualizem o banco de dados. A orientação por assunto de um *data warehouse* proporciona uma visão mais abrangente da organização.

- *Integrado*. A integração está bastante ligada à orientação por assunto. Os *data warehouses* devem colocar os dados de diferentes fontes em um formato consistente. Para isso, devem enfrentar conflitos de nomenclaturas e discrepâncias entre unidades de medida. Espera-se que um *data warehouse* seja totalmente integrado.
- *Variável no tempo (série temporal)*. Um *data warehouse* mantém dados históricos. Os dados não necessariamente mostram o status atual (exceto em sistemas em tempo real). Eles detectam tendências, variações, relações de longo prazo para previsão e comparações, o que leva à tomada de decisões. Há uma qualidade *temporal* para cada *data warehouse*. O tempo é uma dimensão importante à qual todo *data warehouse* deve oferecer suporte. Os dados de análise vindos de diversas fontes contêm diversos pontos de tempo (p. ex., visualizações diárias, semanais, mensais).
- *Não-volátil*. Após os dados serem inseridos em um *data warehouse*, os usuários não podem alterar ou atualizá-los. Os dados obsoletos são descartados e as alterações são registradas como dados novos. Isso permite que o *data warehouse* seja ajustado quase exclusivamente para o acesso a dados.

Características adicionais de um *data warehouse* podem incluir o seguinte:

- *Baseado na Web*. Os *data warehouses* normalmente são criados para apresentar um ambiente informatizado eficiente às aplicações baseadas na Web.
- *Relacional/multidimensional*. Um *data warehouse* usa ou uma estrutura relacional ou uma estrutura multidimensional.
- *Cliente/servidor*. Um *data warehouse* usa a arquitetura cliente/servidor para proporcionar aos usuários acesso fácil.
- *Em tempo real*. Os *data warehouses* mais novos oferecem acesso a dados e recursos de análise em tempo real, ou ativos (consulte Basu: 2003; e Bonde e Kuckuk: 2004).
- *Inclui metadados*. Um *data warehouse* contém metadados (dados sobre dados) acerca de como os dados estão organizados e como usá-los de forma eficiente.

Considerando que o *data warehouse* é um repositório de dados, *data warehousing* é literalmente o processo inteiro (leia Watson: 2002). *Data warehousing* é uma *disciplina* que resulta em aplicações que oferecem capacidade de suporte à decisão, permite acesso imediato às informações de negócios e cria *insights* de negócios. Há três tipos principais de *data warehouses*: *data marts*, *data stores* operacionais (ODS) e *data warehouses* empresariais (EDW). Nossa discussão a seguir incluirá, além destes, os metadados.

Data marts

Um *data warehouse* une bancos de dados de toda uma empresa; já um *data mart* normalmente é menor e se concentra em um assunto ou departamento específico. Um *data mart* é um subconjunto de um *data warehouse*, que normalmente consiste em uma única área temática (p. ex., *marketing*, operações). Um *data mart* pode ser *dependente* ou *independente*. Um *data mart dependente* é um subconjunto criado diretamente a partir do *data warehouse*. Ele tem a vantagem de usar um modelo de dados consistente e apresentar dados de qualidade. Os *data marts* dependentes suportam o conceito de um único modelo de dados em toda a empresa, mas o *data warehouse* deve ser construído antes. Eles garantem que o usuário final visualize a mesma versão de dados acessada por todos os outros usuários do *data warehouse*. O alto custo deste último limita seu uso às grandes empresas. Como alternativa, muitas empresas usam uma versão de *data warehouse* reduzida em custo e escala, denominada *data mart independente*. Um *data mart independente* é um *warehouse* pequeno, projetado para uma unidade estratégica de negócios (UEN) ou um departamento, mas cuja fonte não é um EDW.

Data stores operacionais

Um **data store operacional (ODS)** proporciona uma forma de arquivar informações recentes para consumo (*customer information file* – CIF). Este tipo de banco de dados é constantemente usado como área de preparação temporária de um *data warehouse*. Ao contrário dos conteúdos estáticos de um *data warehouse*, os conteúdos de um ODS são atualizados durante o curso das operações comerciais. Um ODS é usado para decisões de curto prazo que envolvem aplicações vitais, e não para decisões de médio e longo prazo associadas a um EDW. Um ODS é semelhante à memória de curto prazo porque armazena apenas informações muito recentes. Em comparação, um *data warehouse* é como uma memória de longo prazo, pois armazena informações permanentes. Um ODS consolida dados de diversos sistemas-fonte e dá uma visão integrada e quase em tempo real de dados voláteis e correntes. Os processos de ETL (discutidos adiante neste capítulo) de um ODS são idênticos aos de um *data warehouse*. Finalmente, os **oper marts** (consulte Imhoff: 2001) são criados quando os dados operacionais necessitam de análise multidimensional. Os dados de um *oper mart* vêm de um ODS.

Data warehouses empresariais (EDW)

Um **data warehouse empresarial (EDW)** é um *data warehouse* em grande escala usado por toda a empresa no suporte à decisão. O EDW é o tipo de *data warehouse* desenvolvido pela Continental, conforme descrito na vinheta de abertura. Sua natureza de grande escala oferece integração dos dados oriundos de muitas fontes em um formato padronizado, para eficiência das aplicações de BI e suporte à decisão. O EDW é usado para fornecer dados a muitos tipos de DSS, incluindo CRM, gestão da cadeia de fornecimento (SCM), *business performance management* (BPM), monitoramento de atividades de negócios (BAM), gestão do ciclo de vida de produtos (PLM), gestão da receita e, às vezes, até mesmo sistemas de gestão do conhecimento (SGC). Para um exemplo prático, veja o estudo de caso da MindTree Consulting, “Building an Enterprise Data Warehousing for a Major Pharmaceutical Company”, disponível no endereço mindtree.com/clt/cs_dw_pharma.html.

Metadados

Metadados são dados sobre dados (p. ex., consulte Sen: 2004 e Zhao: 2005). Os metadados descrevem a estrutura e alguns significados a respeito dos dados, e, assim, contribuem para que seu uso seja eficiente ou ineficiente. Mehra (2005) indicou que poucas organizações realmente entendem os metadados, e menos ainda são as que entendem como projetar e implementar uma estratégia de metadados. Normalmente, eles são definidos em termos de uso como metadados *técnicos* ou *de negócios*. Outra forma de visualizar os metadados é por *padrão*. Conforme a visualização do padrão, podemos diferenciar *metadados sintáticos* (dados que descrevem a sintaxe dos dados), *metadados estruturais* (dados que descrevem a estrutura dos dados) e *metadados semânticos* (dados que descrevem o significado dos dados em determinado domínio).

Explicaremos a seguir os padrões tradicionais de metadados e idéias sobre como implementar uma estratégia eficaz de metadados por meio de uma abordagem holística à sua integração empresarial. A abordagem inclui ontologia e registros de metadados; integração de informações corporativas (EII); extração, transformação e carga (ETL); e arquitetura orientada a serviços (SOA). Efetividade, extensibilidade, reusabilidade, interoperabilidade, eficiência e desempenho, evolução, poder, flexibilidade, segregação, interface de usuário, versionamento, versatilidade e baixo custo de manutenção são alguns dos principais requisitos para se construir uma empresa bem-sucedida orientada por metadados.

Segundo Kassam (2002), os *metadados de negócios* incluem informações que aumentam nossa compreensão sobre os dados tradicionais (estruturados). O principal objetivo dos metadados é oferecer contexto aos dados relatados, isto é, informações enriquecedoras que levam à criação de conhecimento. Os metadados de negócios, ainda que difíceis de providenciar corretamente,

liberam mais potencial dos dados estruturados. O contexto não precisa ser o mesmo para todos os usuários. De muitas formas, os metadados auxiliam a converter dados e informações em conhecimento. Os metadados formam uma base para uma arquitetura *de metanegócios* (consulte BELL: 2001). Tannenbaum (2002) descreveu como identificar requisitos de metadados. Vaduva e Vetterli (2001) proporcionaram uma visão geral da gestão de metadados para *data warehousing*. Zhao (2005) descreveu cinco níveis de maturidade da gestão de metadados: 1) *ad hoc*, 2) descoberta, 3) gerenciada, 4) otimizada e 5) automatizada. Esses níveis ajudam a compreender onde uma organização se encontra em termos de como usa e se faz bom uso de seus metadados.

O projeto, a criação e o uso dos metadados – dados descritivos ou resumidos sobre dados – e os padrões que os acompanham podem acarretar questões éticas. Há considerações éticas envolvidas na coleta e posse das informações contidas nos metadados, inclusive questões intelectuais e de privacidade que surgem nos estágios de projeto, coleta e disseminação (saiba mais em Brody: 2003).

Questões de revisão da Seção 2.2

1. O que é um *data warehouse*?
2. Como um *data warehouse* se diferencia de um banco de dados?
3. O que é um ODS?
4. Explique a diferença entre um *data mart*, um ODS e um EDW.
5. Explique a importância dos metadados.

2.3

VISÃO GERAL DO PROCESSO DE DATA WAREHOUSING

Organizações públicas e privadas constantemente coletam dados, informações e conhecimento em níveis cada vez maiores, e os armazenam em sistemas informatizados. Manter e usar esses dados e informações se torna extremamente complexo, principalmente quando surgem questões de escalabilidade. Além disso, o número de usuários que precisam acessar as informações continua aumentando, como resultado da maior confiabilidade e disponibilidade do acesso à rede, sobretudo à Internet. O trabalho com múltiplos bancos de dados, integrados ou não a um *data warehouse*, tornou-se uma tarefa extremamente difícil e que exige um grande conhecimento especializado; porém, ele pode oferecer benefícios imensos que superam em muito seu custo (leia a vinheta de abertura e o Caso de aplicação 2.1).

CASO DE APLICAÇÃO 2.1

O *data warehousing* apóia a estratégia corporativa da First American Corporation

A First American Corporation alterou sua estratégia corporativa de uma abordagem bancária tradicional para uma centrada em CRM. Com isso, foi possível a ela deixar de ser uma empresa que perdeu US\$ 60 milhões, em 1990, para transformar-se em uma líder inovadora em serviços financeiros, uma década depois. A implementação bem-sucedida desta estratégia não seria possível sem o *data warehouse* VISION, que armazena informações sobre comportamento dos clientes, como produtos usados, prefe-

rências de compras e posições de valor do cliente. O VISION oferece:

- Identificação dos principais 20% de clientes rentáveis
- Identificação de 40 a 50% dos clientes não-rentáveis
- Estratégias de retenção
- Canais de distribuição de menor custo
- Estratégias para ampliar as relações com clientes
- Fluxos de informação reprojatados

O acesso à informação por meio de um *data warehouse* pode permitir alterações evolucionárias e revolucionárias. A First American Corporation obteve uma alteração revolucionária, passando a ser uma das “16 principais” entre as corporações de serviços financeiros.

Fontes: adaptado de B.L. Cooper, H.J. Watson, B.H. Wixom, and D.L. Goodhue, “Data Warehousing Supports Corporate Strategy at First American Corporation”, *MIS Quarterly*, Vol. 24, No. 4, 2000, pp. 547–567; e B.L. Cooper, H.J. Watson, B.H. Wixom, e D.L. Goodhue, *Data Warehousing Supports Corporate Strategy at First American Corporation*, SIM International Conference, Atlanta, August, 15–19, 1999.

Muitas organizações necessitam criar *data warehouses* – armazéns gigantescos com dados em série temporal para o suporte à decisão. Os dados são importados de vários recursos internos e externos, e são limpos e organizados de forma coerente com as necessidades da organização. Após preencher o *data warehouse* com os dados, *data marts* podem ser carregados para uma área ou departamento específico. Os *data marts* podem também ser criados primeiro, conforme necessário, e depois integrados ao EDW. No entanto, com frequência os *data marts* não são desenvolvidos, e os dados são simplesmente carregados nos PCs ou mantidos em seu estado original para serem manipulados diretamente com uso de ferramentas de BI.

Na Figura 2.1, mostramos o conceito de *data warehouse*. Estes são os principais componentes de um processo de *data warehousing*:

- **Fontes de dados.** Os dados são obtidos de múltiplos sistemas operacionais independentes “legados” e, possivelmente, de fornecedores externos de dados (como o U.S. Census). Os dados podem também vir de um sistema de processamento de transações *online* (OLTP) ou de sistemas integrados de gestão (ERP). Dados da Web na forma de *Web logs* também alimentam um *data warehouse*.

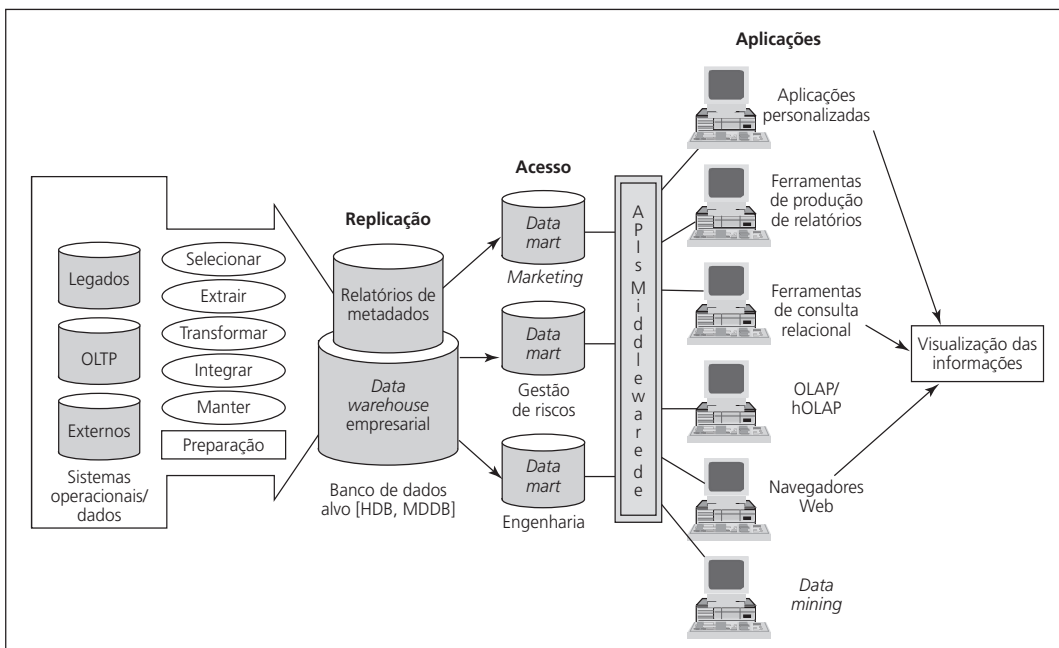


Figura 2.1 Estrutura e visualizações do *data warehouse*.

- *Extração de dados.* Os dados são extraídos com um *software* personalizado ou comercial chamado ETL.
- *Carregamento de dados.* Carregam-se os dados em uma área de preparação, na qual eles são transformados e limpos. Os dados, então, ficam prontos para o carregamento no *data warehouse*.
- *Banco de dados abrangente.* Basicamente, este é o EDW que suporta todas as análises de decisões fornecendo informações relevantes resumidas e detalhadas, originais de muitas fontes diferentes.
- *Metadados.* Os metadados são mantidos para que sejam acessados pelo pessoal de TI e pelos usuários. Eles incluem programas sobre dados e regras para organizar resumos de dados que sejam fáceis de indexar e buscar, principalmente com ferramentas da Web.
- *Ferramentas de middleware.* As ferramentas de *middleware* permitem o acesso ao *data warehouse*. Usuários avançados, como analistas, podem escrever suas próprias consultas SQL. Outros podem empregar um ambiente gerenciado de consultas, como Business Objects, para acessar os dados. Há muitas aplicações *front-end* que os usuários de negócios podem usar para interagir com os dados armazenados nos repositórios, incluindo *data mining*, OLAP e ferramentas de relatórios e de visualização de dados.

Questões de revisão da Seção 2.3

1. Descreva o processo de *data warehousing*.
2. Descreva os principais componentes de um *data warehouse*.
3. Identifique o papel das ferramentas de *middleware*.

2.4

ARQUITETURAS DE DATA WAREHOUSING

Há algumas arquiteturas básicas de *data warehousing*. As arquiteturas de duas e de três camadas são comuns (veja as Figuras 2.2 e 2.3), mas, às vezes, há apenas uma camada. Hoffer et al. (2007) as distinguem dividindo o *data warehouse* em três partes:

1. O próprio *data warehouse*, que contém os dados e o *software* associados.
2. *Software* de aquisição de dados (retaguarda), que extrai dados de sistemas legados e fontes externas, os consolida e resume, e depois os carrega no *data warehouse*.
3. *Software* cliente (*front-end*), que permite aos usuários acessar e analisar dados a partir do *warehouse* (um mecanismo de DSS/BI/análise de negócios [BA]).

Na arquitetura de três camadas, os sistemas operacionais contêm os dados e o *software* para aquisição em uma camada (o servidor), o *data warehouse* em outra camada, e a terceira camada inclui o mecanismo de DSS/BI/BA (o servidor de aplicação) e o cliente (veja a Figura 2.2). Os dados do *warehouse* são processados duas vezes e depositados em um banco de dados multidimensional adicional, organizado para facilitar a análise e apresentação multidimensional, ou replicados em *data marts*. A vantagem da arquitetura de três camadas é sua separação das funções do *data warehouse*, que elimina as limitações de recursos e possibilita a fácil criação de *data marts*.

Em uma arquitetura de duas camadas, o mecanismo de DSS é executado fisicamente na mesma plataforma de *hardware* que o *data warehouse* (veja a Figura 2.3). Portanto, é mais econômico que a arquitetura de três camadas. A arquitetura de duas camadas pode apresentar problemas de desempenho, caso haja *data warehouses* grandes que trabalhem com aplicações de uso intenso de dados no suporte à decisão.

A sabedoria popular, em boa parte, assume uma abordagem absolutista, afirmando que uma solução é melhor que a outra, independentemente das circunstâncias e necessidades exclusivas da organização. Para complicar ainda mais essas decisões de arquitetura, muitos consultores e fornecedores de *software* se concentram em uma parte da arquitetura. Portanto, acabam limitando

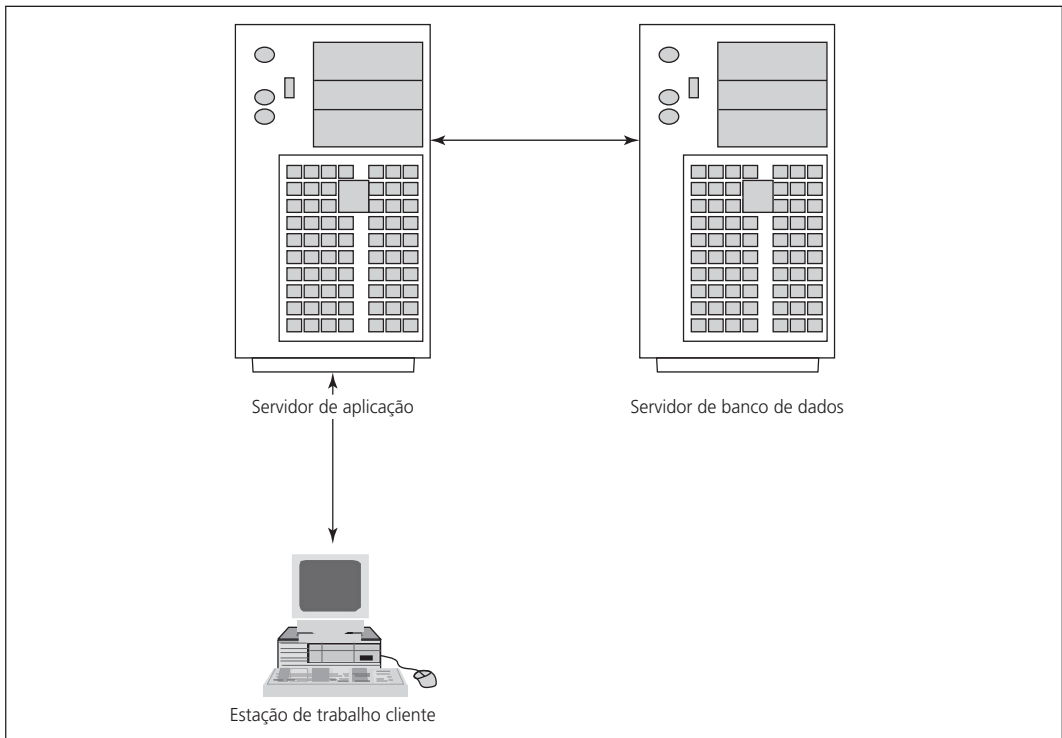


Figura 2.2 Arquitetura de um *data warehouse* de três camadas.

sua capacidade e motivação para auxiliar uma organização a ver as opções com base em suas necessidades. Entretanto, esses aspectos estão sendo questionados e analisados. Por exemplo, Ball (2005) forneceu critérios de decisão para as organizações que têm planos de implementar uma aplicação de BI e já determinaram suas necessidades de *data marts* multidimensionais, mas que precisam de ajuda para decidir a arquitetura em camadas apropriada. Os critérios do autor giram em torno de necessidades de previsão para obtenção de espaço e velocidade de acesso (para mais detalhes, consulte Ball: 2005).

O *data warehousing* e a Internet são duas tecnologias essenciais que oferecem soluções importantes à gestão de dados corporativos. A integração dessas duas tecnologias produz o *data warehousing* baseado na Web. Na Figura 2.4, mostramos a sua arquitetura. Ela possui três camadas e inclui PC cliente, servidor da Web e servidor de aplicação. No lado do cliente, o usuário necessita de uma conexão à Internet e um navegador da Web (de preferência, que permita Java) ligado à interface gráfica familiar ao usuário (GUI). A *Internet/intranet/extranet* é o meio de comunicação entre cliente e servidores. No lado do servidor, um servidor da Web é usado para administrar os fluxos de entrada e saída de informações entre cliente e servidor. Ele é auxiliado por um *data warehouse* e por um servidor de aplicação. O *data warehousing* baseado na Web oferece algumas vantagens atraentes, que incluem facilidade de acesso, independência de plataforma e custo mais baixo.

O Vanguard Group adotou uma arquitetura de três camadas baseada na Web, para que a arquitetura de sua empresa integrasse todos os dados e oferecesse aos clientes as mesmas visualizações de dados que tinham os usuários internos (Dragoon: 2003). Da mesma forma, o Hilton migrou todos os seus sistemas cliente/servidor independentes para um *data warehouse* de três camadas, usando um sistema empresarial com projeto da Web. A mudança incluiu um investi-

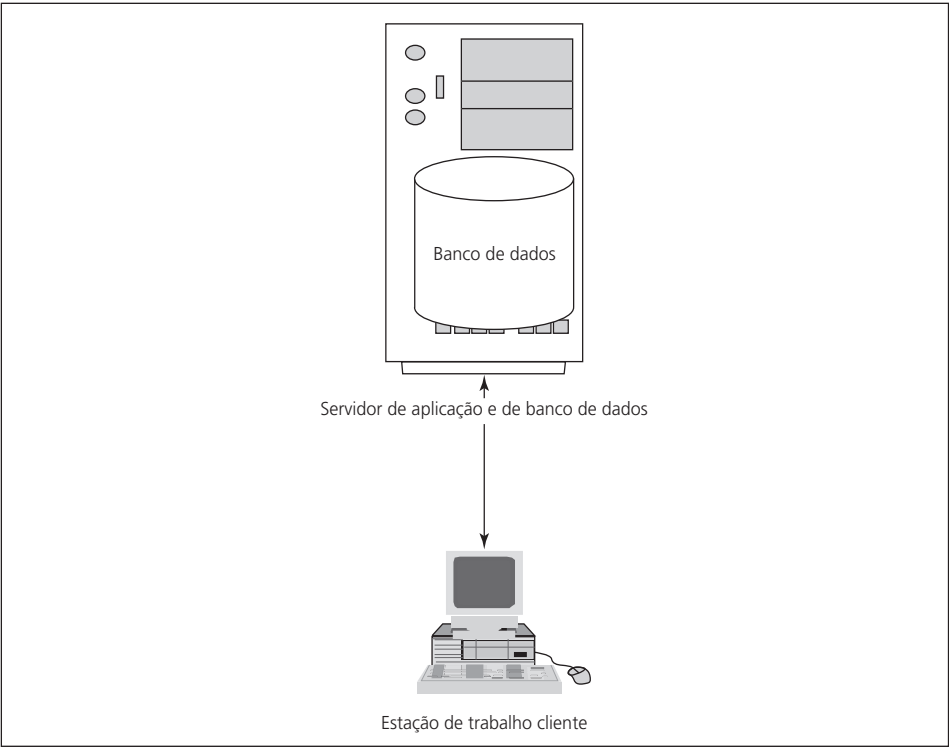


Figura 2.3 Arquitetura de um *data warehouse* de duas camadas.

mento de US\$ 3,8 milhões (excluindo a mão-de-obra) e afetou 1.500 usuários. Ela aumentou a eficiência de processamento (velocidade) em seis vezes. Quando foi implementada, o Hilton esperava economizar entre US\$ 4,5 e 5 milhões ao ano. Finalmente, a corporação fez experiências com a tecnologia de *clustering* (computação paralela) da Dell para aprimorar a escalabilidade e a velocidade (veja Anthes: 2003).

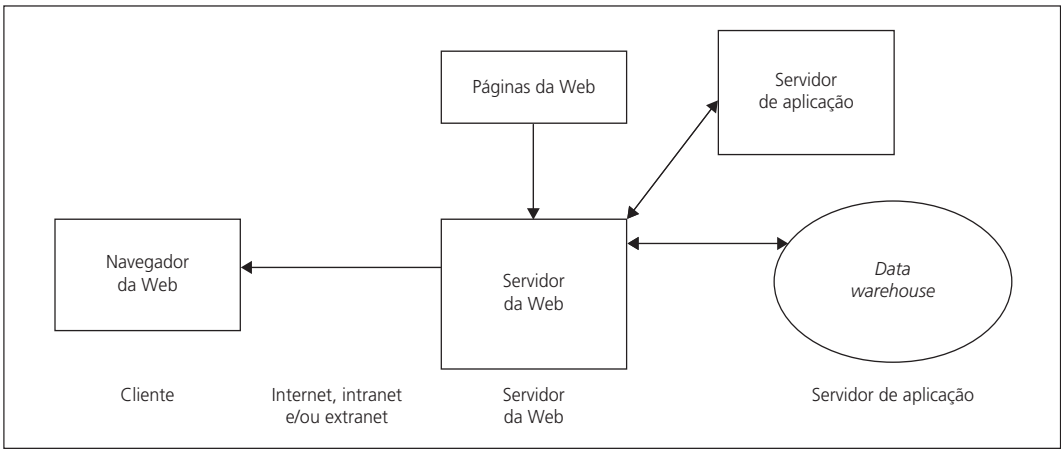


Figura 2.4 Arquitetura do *data warehousing* baseado na Web.

As arquiteturas de *data warehousing* da Web têm estrutura semelhante a outras arquiteturas de *data warehousing* e exigem uma escolha de projeto para hospedar o *warehouse* na Web com o servidor de transações ou como servidor(es) separado(s). A velocidade de carregamento das páginas é uma consideração importante ao se projetar aplicações baseadas na Web; portanto, se deve planejar cuidadosamente a capacidade do servidor.

Há algumas questões a serem consideradas ao decidir que arquitetura usar. Eis algumas delas:

- *Qual sistema de gerenciamento de banco de dados (SGBD) usar?* A maioria dos *data warehouses* é construída usando sistemas de gerenciamento de banco de dados relacional (SGBDR). Oracle (Oracle Corporation, www.oracle.com), SQL Server (Microsoft Corporation, microsoft.com/sql/) e DB2 (IBM Corporation, www-306.ibm.com/software/data/db2/) são os mais usados. Todos esses produtos suportam arquiteturas cliente/servidor e baseada na Web.
- *O processamento paralelo e/ou o particionamento serão usados?* O processamento paralelo permite que múltiplas CPUs processem solicitações de consultas ao *data warehouse* simultaneamente, além de oferecer escalabilidade. Os projetistas do *data warehouse* precisam decidir se as tabelas de bancos de dados serão particionadas (divididas em tabelas menores), para maior eficiência de acesso, e quais serão os critérios. Esta é uma consideração importante, que se faz necessária pelas grandes quantidades de dados contidas em um *data warehouse* típico. A Teradata (teradata.com) adotou esta abordagem com sucesso.
- *Serão usadas ferramentas de migração de dados para carregar o data warehouse?*
- *Que ferramentas serão usadas para dar suporte à recuperação e análise de dados?*

Arquiteturas alternativas

Os pontos de vistas de projeto arquitetônico do *data warehouse* podem ser classificados, de maneira geral, em projeto de *data warehouse* para toda a empresa e projeto de *data mart*. Na Figura 2.5 (a-e), mostramos algumas alternativas para os tipos básicos de projeto arquitetônico, incluindo uma arquitetura em estrela, um *warehouse* empresarial com ODS (isto é, suporte ao acesso em tempo real) e uma arquitetura de EDW distribuída. Sen e Sinha (2005) analisaram 15 metodologias diferentes de *data warehousing*. As fontes dessas metodologias são classificadas em três am-

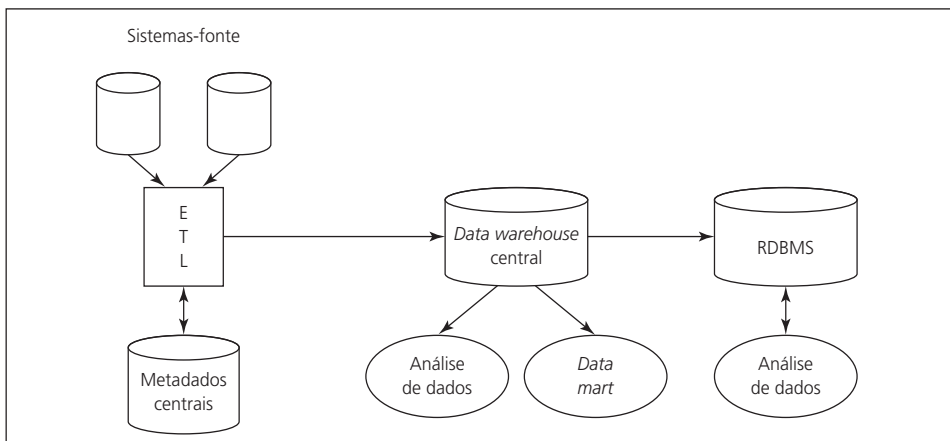


Figura 2.5 Arquiteturas alternativas de data warehouse.

2.5a Arquitetura de *data warehousing* empresarial.

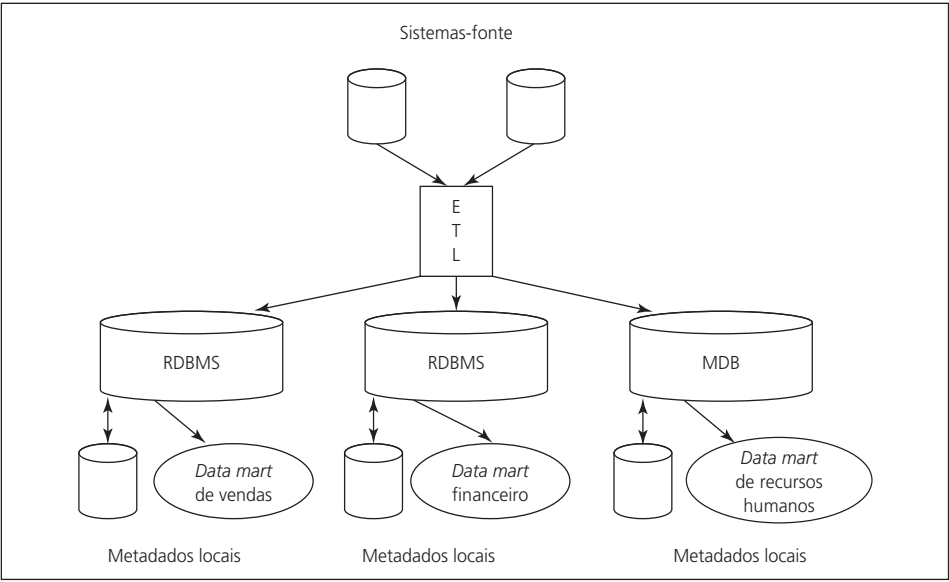


Figura 2.5b Arquitetura de data mart.

plas categorias: fornecedores da tecnologia específica, fornecedores de infra-estruturas e empresas de modelagem de informações. Consulte Sen e Sinha (2005) para maiores detalhes.

A literatura sobre *data warehousing* oferece mais discussões sobre diversas arquiteturas, como *data marts* independentes, arquitetura de barramento de *data mart* com *data marts* dimensionais vinculados, e *data marts* federados (consulte Ariyachandra and Watson: 2005, 2006a, 2006b),

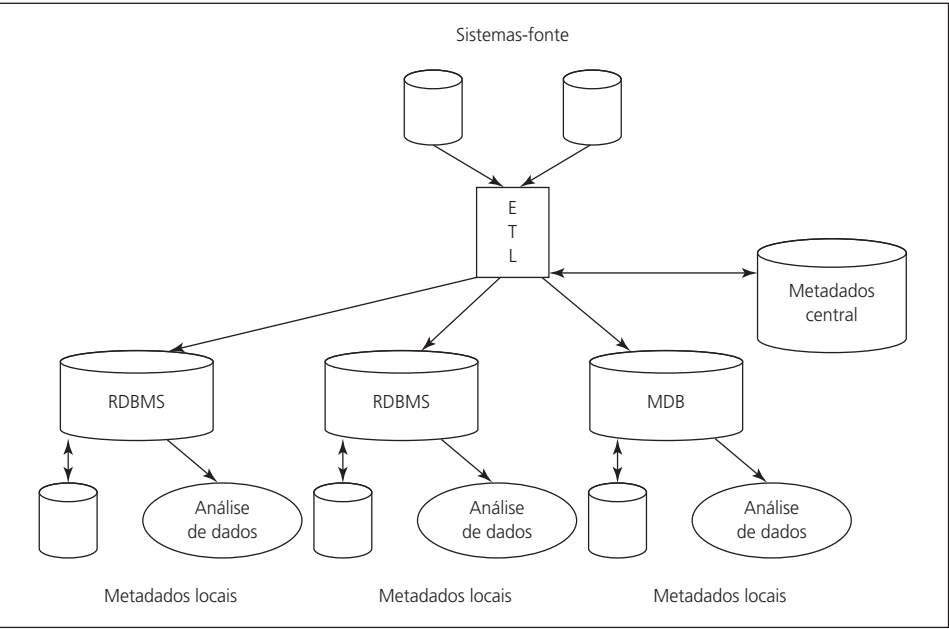


Figura 2.5c Arquitetura em estrela de data mart.

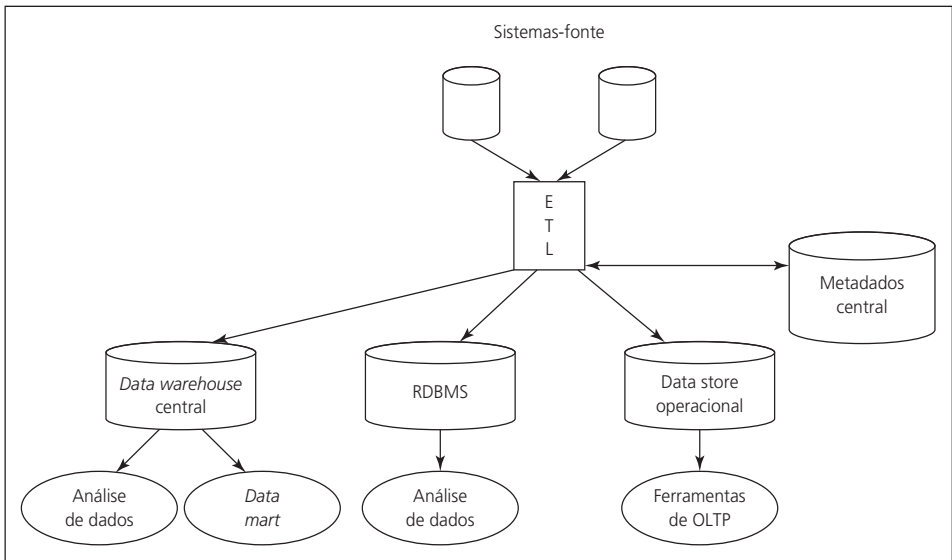


Figura 2.5d Warehouse empresarial e data store operacional.

veja a Figura 2.6. Nos *data marts* independentes, os depósitos são desenvolvidos para operar de forma autônoma uns dos outros. Assim, eles têm definições inconsistentes de dados e diferentes medidas e dimensões, o que dificulta a análise de seus dados (ou seja, é difícil, se não impossível, ter “a versão única da verdade”). Na arquitetura em estrela, a atenção se volta para a construção de

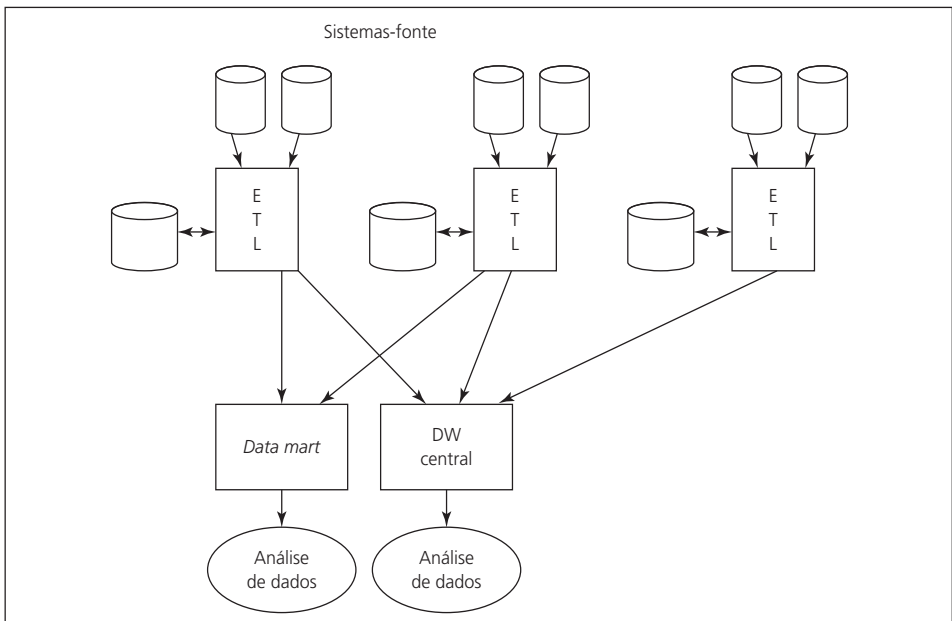


Figura 2.5e Arquitetura de data warehouse distribuída.

Fonte: adaptado de A. Sen and P. Sinha, “A Comparison of Data Warehousing Methodologies”, *Communications of the ACM*, Vol. 48, No. 3, 2005, pp. 78–84; e T. Ariyachandra e H. Watson, “Which Data Warehouse Architecture Is Most Successful?” *Business Intelligence Journal*, Vol. 11, No. 1, First Quarter 2006, pp. 4–6.

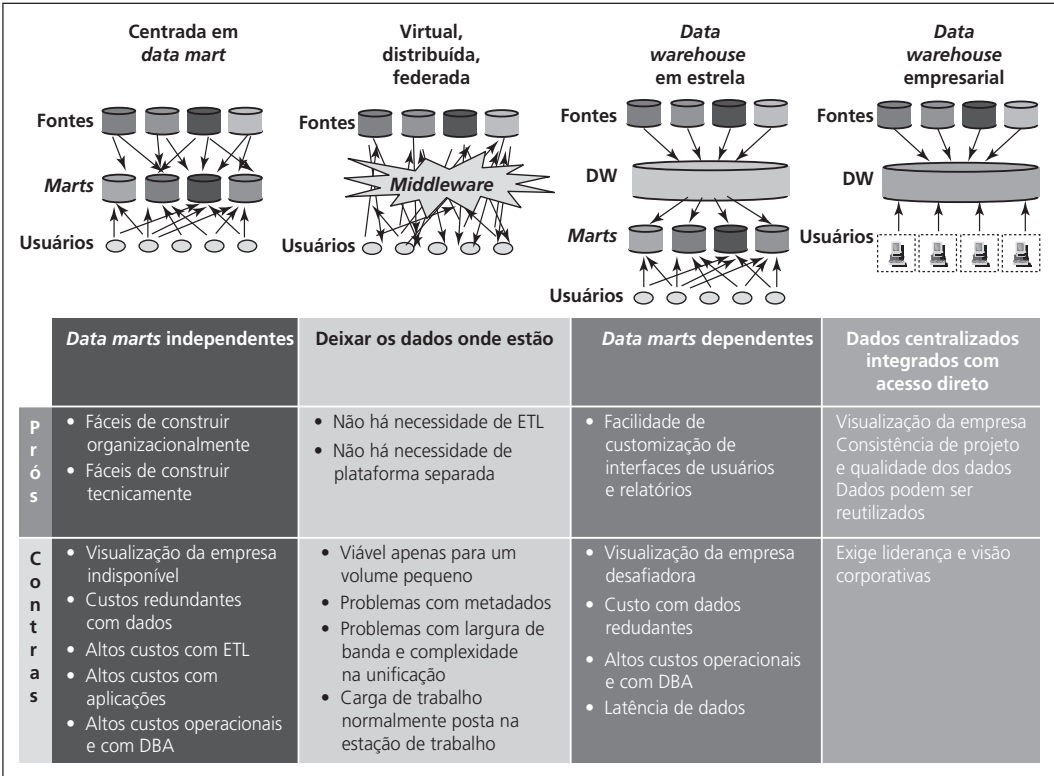


Figura 2.6 Arquiteturas alternativas para empreendimentos de data warehousing.

Fonte: W. Eckerson, "Four Ways to Build a Data Warehouse", *What Works: Best Practices in Business Intelligence and Data Warehousing*, Vol. 15, The Data Warehousing Institute, Chatsworth, CA, June 2003, pp. 46–49. Usado mediante permissão.

uma infra-estrutura escalonável e passível de manutenção; ela é desenvolvida de forma iterativa, área temática por área temática, e são desenvolvidos *data marts* dependentes. Um *data warehouse* centralizado é semelhante à arquitetura em estrela; porém, não há *data marts* dependentes. A arquitetura de *data warehouses* centrais, defendida principalmente pela Teradata Corp., aconselha o uso de *data warehouses* sem nenhum *data mart* (veja a Figura 2.7). Esta abordagem centralizada oferece aos usuários acesso a todos os dados no *data warehouse*, em vez de limitá-los aos *data marts*. Além do mais, ela reduz a quantidade de dados que a equipe técnica deve transferir ou alterar, simplificando a gestão e a administração dos dados.

A *abordagem federada* é uma concessão às forças naturais que minam os melhores planos de desenvolvimento de um sistema perfeito. Ela usa todos os meios possíveis para integrar recursos analíticos de múltiplas fontes, para atender às necessidades em alteração ou às condições de negócios. Basicamente, a abordagem federada envolve integrar sistemas díspares. Em uma arquitetura federada, as estruturas existentes de suporte à decisão são mantidas no lugar, e os dados são acessados a partir destas fontes, conforme necessário. A abordagem federada recebe suporte dos fornecedores de *middleware* que propõem recursos de consulta distribuída e unificação. Essas ferramentas baseadas em Extensible Markup Language (XML) oferecem uma visualização global das fontes de dados distribuídas, incluindo *data warehouses*, *data marts*, *websites*, documentos e sistemas operacionais. Quando os usuários escolhem consultar objetos desta visualização e pressionam o botão de envio, a ferramenta automaticamente consulta as fontes distribuídas, unifica os resultados e os apresenta aos usuários. Devido a problemas de desempenho e qualidade dos dados, a maioria dos especialistas concorda que

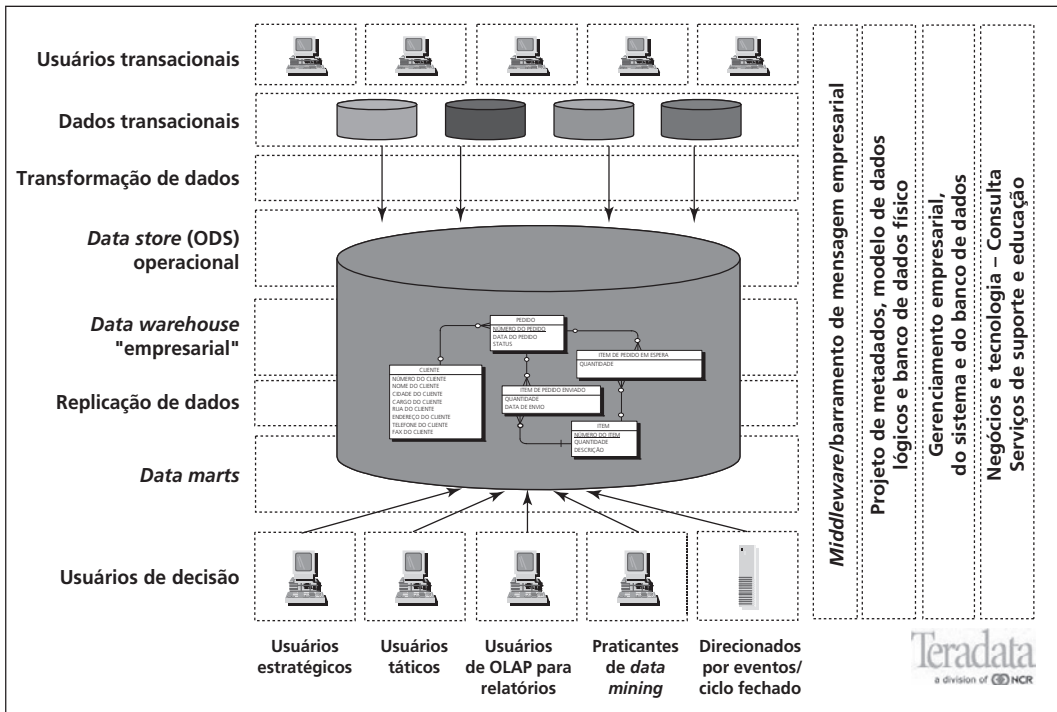


Figura 2.7 Data warehouse empresarial da Teradata Corp.

Fonte: Teradata Corporation (teradata.com). Usado mediante permissão.

as abordagens federadas funcionam bem para complementar os *data warehouses*, e não para substituí-los (leia Eckerson: 2005).

Ariyachandra e Watson (2005) identificaram 10 fatores que possivelmente afetam a decisão de seleção da arquitetura:

1. Interdependência de informações entre as unidades da organização
2. Necessidades de informações da alta administração
3. Necessidade urgente de um *data warehouse*
4. Natureza das tarefas do usuário final
5. Limitação de recursos
6. Visão estratégica do *data warehouse* antes da implementação
7. Compatibilidade com os sistemas existentes
8. Observação da capacidade da equipe interna de TI
9. Questões técnicas
10. Fatores sociais/políticos

Esses fatores são semelhantes a muitos fatores de sucesso descritos na literatura para projetos de sistemas de informação e projetos de DSS e BI. Mais do que proporcionarem tecnologia viavelmente pronta para uso, as questões técnicas são importantes, entretanto, não são tão importantes quanto as questões comportamentais, como a satisfação das necessidades de informação da alta administração e o envolvimento dos usuários no processo de desenvolvimento (um fator social/político). Cada arquitetura de *data warehousing* tem aplicações específicas para as quais sua eficiência é maior – ou menor –, e assim oferece os benefícios máximos para a organi-

zação. Porém, no geral, a estrutura de *data mart* parece ser a menos eficiente na prática (consulte Ariyachandra and Watson: 2006b). Leia Ariyachandra e Watson (2006a) para maiores detalhes.

Questões de revisão da Seção 2.4

1. Quais são as principais semelhanças e diferenças entre uma arquitetura de duas camadas e uma arquitetura de três camadas?
2. Como a Web influenciou os projetos de *data warehouse*?
3. Liste as arquiteturas alternativas de *data warehousing* discutidas nesta seção.
4. Que questões devem ser consideradas ao se decidir qual arquitetura usar no desenvolvimento de um *data warehouse*? Liste os 10 fatores mais importantes.

2.5

INTEGRAÇÃO DE DADOS E PROCESSOS DE EXTRAÇÃO, TRANSFORMAÇÃO E CARGA (ETL)

As pressões competitivas globais, a demanda por ROI, as investigações de gerentes e investidores e as regulamentações governamentais forçam os diretores de empresas a repensar a forma como integram e gerenciam seus negócios. Um tomador de decisão normalmente precisa de acesso a múltiplas fontes de dados que devem ser integradas. Antes dos *data warehouses*, dos *data marts* e dos *software* de BI, oferecer acesso às fontes de dados era um processo trabalhoso e de grande porte. Mesmo com as modernas ferramentas de gestão de dados baseadas na Web, reconhecer quais dados acessar e fornecê-los ao tomador de decisão é uma tarefa complicada, que exige especialistas em bancos de dados. À medida que crescem os *data warehouses*, aumentam as questões sobre integração de dados.

As necessidades de BA continuam a evoluir. Fusões e aquisições, legislações regulamentares e a introdução de novos canais podem provocar mudanças nas exigências de BI. Além de dados históricos, limpos, consolidados e pontuais no tempo, os usuários de negócios cada vez mais exigem acesso a dados em tempo real, não-estruturados e/ou remotos. E tudo deve ser integrado aos conteúdos de um *data warehouse* existente (consulte Devlin: 2003). Ademais, o acesso via PDAs e por meio de reconhecimento e síntese de voz torna-se mais e mais corriqueiro, o que complica as questões de integração (leia Edwards: 2003). Muitos projetos de integração envolvem sistemas de toda a empresa. Orovic (2003) apresentou uma lista de verificação de o que funciona e o que não funciona quando se tenta fazer um projeto como esse. É difícil integrar adequadamente dados de vários bancos e outras fontes distintas. Porém, quando não realizada adequadamente, essa tarefa pode ser desastrosa para os sistemas de toda a empresa, como CRM, ERP e projetos de sistemas da cadeia de fornecimento. Veja mais em Nash (2002), Dasu e Johnson (2003).

INTEGRAÇÃO DE DADOS

A **integração de dados** compreende três grandes processos que, quando implementados corretamente, permitem que os dados sejam acessados por e disponibilizados a uma gama de ferramentas de ETL e análise e ao ambiente de *data warehousing*. Os processos são: acesso aos dados (a capacidade de acessar e extrair dados de qualquer fonte), federação de dados (a integração das visualizações de negócios em diversos *data stores*) e captura de alterações (com base na identificação, captura e entrega das alterações feitas nas fontes de dados da empresa). Leia os detalhes em Sapir (2005). Veja no Caso de aplicação 2.2 um exemplo de como o Bank of America se beneficia da implementação de um *data warehouse* que integra dados de muitas fontes. Alguns fornecedores, como o SAS Institute, Inc., desenvolveram fortes ferramentas de integração de dados. O servidor corporativo de integração de dados do SAS inclui ferramentas de integração de dados de clientes que melhoram a qualidade dos dados no processo de integração. O Oracle Business Intelligence Suite também auxilia na integração dos dados.

CASO DE APLICAÇÃO 2.2

O data warehouse integrado e premiado do Bank of America

Em 2003, o Bank of America venceu o prêmio Best Practices and Leadership (melhores práticas e liderança) do Data Warehousing Institute. Uma das maiores redes de serviços financeiros nos Estados Unidos, o Bank of America chegou a uma economia operacional significativa através da *integração* dos seus *data warehouses*. O Teradata Warehouse é a plataforma para seu EDW integrado. O *data warehouse* auxilia os tomadores de decisão a:

- Manter a privacidade do cliente.
- Fazer um bom uso das informações do cliente para desenvolver produtos e identificar tendências.
- Antecipar as necessidades do cliente, ocasionando melhorias nos serviços ao cliente e nas vendas.

- Diminuir custos, melhorar o uso e o desempenho e reagir rapidamente às mudanças das demandas comerciais.
- Tomar decisões melhores e mais rápidas.

Fontes: M. Edwards, “2003 Best Practices Awards Winners: Innovators in Business Intelligence and Data Warehousing”, *Business Intelligence Journal*, Vol. 8, No. 4, Fall 2003, pp. 57–64; NCR, *Bank of America Expands Teradata Data Warehouse System*, October 6, 2005, ncr.com/en/media_information/2005/oct/pr100605a.htm (acessado em abril de 2006); and Teradata, *Bank of America Expands Teradata Data Warehouse System*, October 6, 2005, teradata.com/t/page/141826/index.html (acessado em abril de 2006).

Um grande objetivo do *data warehouse* é integrar dados de múltiplos sistemas. Várias tecnologias de integração permitem atualmente a integração de dados e metadados:

- Integração de aplicações corporativas (EAI)
- Arquitetura orientada a serviços (SOA)
- Integração de informações corporativas (EII)
- Extração, transformação e carga (ETL)

A **integração de aplicações corporativas (EAI)** proporciona um veículo para transportar dados dos sistemas-fonte ao *data warehouse*. Ela envolve a integração das funcionalidades da aplicação e se concentra no compartilhamento de funcionalidades (e não de dados) por todos os sistemas, o que permite flexibilidade e reutilização. Tradicionalmente, as soluções de EAI se concentravam em permitir a reutilização da aplicação, no nível da interface de programação de aplicações (API). Recentemente, obtém-se a EAI utilizando serviços de SOA de granularidade grossa (uma coleção de processos ou funções de negócios) que são bem definidos e documentados. Usar os *Web services* é uma forma especializada de implementar uma SOA. A EAI pode ser utilizada para facilitar a aquisição de dados diretamente em um *data warehouse* quase em tempo real, ou para oferecer decisões aos sistemas de OLTP. Há muitas abordagens e ferramentas diferentes para a implementação de EAI.

A **integração de informações corporativas (EII)** é um espaço de ferramentas em desenvolvimento que promete integração de dados em tempo real a partir de diversas fontes, como bancos de dados relacionais, *Web services* e bancos de dados multidimensionais. É um mecanismo que extrai dados de sistemas-fonte para atender a um pedido de informação. As ferramentas de EII usam metadados predefinidos para preencher visualizações, que fazem os dados integrados parecerem relacionais para os usuários finais. O XML pode ser o aspecto mais importante da EII, pois permite que os dados recebam *tags* no momento da criação ou mais tarde. Essas *tags* podem ser ampliadas e modificadas para acomodar praticamente qualquer área do conhecimento (leia Kay: 2005).

A integração física de dados costumava ser o principal mecanismo para criar uma visualização integrada com *data warehouses* e *data marts*. Com o advento das ferramentas de EII (veja Kay: 2005), novos padrões de integração virtual de dados são viáveis. Manglik e Mehra (2005) discutiram as vantagens e restrições dos novos padrões de integração de dados, que podem expandir as tradicionais metodologias físicas e apresentar uma visualização abrangente à empresa.

A seguir, nos voltaremos à abordagem de carregamento dos dados no *warehouse*: ETL.

Extração, transformação e carga

No coração da parte técnica do processo de *data warehousing* estão **extração, transformação e carga (ETL)**. As tecnologias de ETL, que já existem há algum tempo, são providenciais para o processo e uso de *data warehouses*. O processo de ETL é um componente integral de qualquer projeto centrado em dados. Os gerentes de TI constantemente enfrentam desafios, pois os processos de ETL costumam consumir 70% do tempo em um projeto centrado em dados.

O processo de ETL consiste em *extração* (leitura dos dados de um ou mais bancos de dados), *transformação* (conversão dos dados extraídos de sua forma anterior na forma em que precisam estar, para que sejam colocados em um *data warehouse* ou apenas em outro banco de dados) e *carga* (colocação dos dados no *data warehouse*). A transformação ocorre com o uso de regras ou tabelas de busca ou com a combinação dos dados com outros dados. As três funções de banco de dados são integradas em uma ferramenta para extrair dados de um ou mais bancos e colocá-los em outro banco de dados ou *data warehouse* consolidados.

As ferramentas de ETL também transportam dados entre fontes e alvos, documentam como os elementos de dados (p. ex., metadados) mudam conforme se movimentam entre fonte e alvo, trocam metadados com outras aplicações conforme necessário, e administram todos os processos e operações de tempo de execução (como programação, gerenciamento de erros, registros de auditorias, estatísticas). A ETL é extremamente importante na integração de dados e também no *data warehousing*. O objetivo do processo de ETL é carregar dados integrados e limpos no *warehouse*. Os dados usados nestes processos podem ser oriundos de qualquer fonte: uma aplicação de *main-frame*, uma aplicação de ERP, uma ferramenta de CRM, um arquivo texto, uma planilha do Excel ou até uma fila de mensagens. Na Figura 2.8, apresentamos um esboço do processo de ETL.

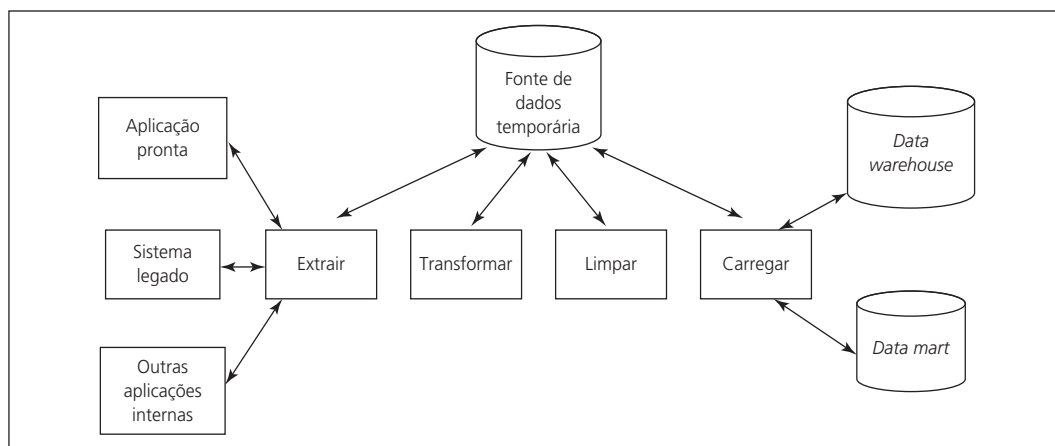


Figura 2.8 O processo de ETL.

Fonte: adaptado de M.L. Songini, "ETL Quickstudy", *Computer World*, February 2, 2004; e T.Ariyachandra e H.Watson, "Which Data Warehouse Architecture Is Most Successful?" *Business Intelligence Journal*, Vol. 11, No. 1, First Quarter 2006, pp. 4-6.

O processo de migração de dados para um *data warehouse* envolve a extração de dados de todas as fontes relevantes. As fontes de dados consistem em arquivos extraídos de bancos de dados OLTP, planilhas, bancos de dados pessoais (p. ex., Microsoft Access) ou arquivos externos. Todos os arquivos de entrada são gravados em um conjunto de tabelas temporárias, criadas para facilitar o processo de carga. Um *data warehouse* contém inúmeras regras de negócios que definem questões como a forma de utilização dos dados, regras de resumo, padronização dos atributos codificados e regras de cálculo. Qualquer problema na qualidade dos dados pertencente aos arquivos-fonte precisa ser corrigido antes que os dados sejam carregados no *data warehouse*. Um dos benefícios de um *data warehouse* bem projetado é que essas regras podem ser armazenadas em um repositório de metadados e aplicadas ao *warehouse* de forma central. Isso difere de uma abordagem de OLTP, em que normalmente os dados e as regras de negócios estão espalhados por todo o sistema. O processo de carregar dados em um *data warehouse* pode ser realizado por meio de ferramentas de transformação de dados que fornecem uma GUI para auxiliar no desenvolvimento e manutenção das regras de negócios. Esse processo pode ainda ser executado por meios mais tradicionais, como desenvolvimento de programas ou utilitários para carregar o *data warehouse*, usando linguagens de programação como PL/SQL, C++ ou .Net Framework. Esta não é uma decisão fácil para as organizações. Diversas questões afetam a resolução de uma empresa que está por adquirir ferramentas de transformação de dados ou construir por si própria o processo de transformação:

- As ferramentas de transformação de dados são caras.
- As ferramentas de transformação de dados têm uma longa curva de aprendizado.
- É difícil medir o desempenho da organização de TI até que ela aprenda a usar as ferramentas de transformação de dados.

A longo prazo, uma abordagem de ferramenta de transformação deve simplificar a manutenção do *data warehouse* de uma organização. As ferramentas de transformação também podem ser eficientes para detectar e limpar (ou seja, remover anomalias nos dados). Ferramentas OLAP e de *data mining* dependem do grau de transformação dos dados.

Como exemplo de uma ETL eficiente, a Motorola Inc. usa ETL para abastecer seus *data warehouses*. A Motorola coleta informações de 30 diferentes sistemas de aquisição e os envia para seu *data warehouse global* de SCM para uma análise dos gastos agregados da empresa (veja SONGINI: 2004).

Solomon (2005) classificou as tecnologias de ETL em quatro categorias: sofisticadas, capacitadoras, simples e rudimentares. Geralmente se reconhece que as ferramentas sofisticadas terão como resultado um processo de ETL mais bem documentado e gerenciado com mais precisão conforme o projeto de *data warehouse* se desenvolve.

Ainda que seja possível os programadores desenvolverem *software* para a ETL, é mais simples usar uma ferramenta existente. Eis alguns dos critérios importantes (BROWN: 2004) para selecionar uma ferramenta de ETL que inclua:

- Capacidade de ler de e gravar em um número ilimitado de arquiteturas de fontes de dados.
- Captura e entrega automática de metadados.
- Histórico de conformidade com padrões abertos.
- Interface fácil de usar para o desenvolvedor e o usuário funcional.

A extensa realização de ETL pode ser um sinal de dados mal gerenciados e de uma falta básica de estratégia coerente de sua gestão. Karacsony (2006) indicou que há uma correlação direta entre a proporção de dados redundantes e o número de processos de ETL. Quando eles são gerenciados corretamente como um ativo da empresa, os esforços de ETL são reduzidos consideravelmente e os dados redundantes são completamente eliminados. Isso resulta em grandes economias com manutenção e maior eficiência em novos desenvolvimentos, enquanto melhora também a qualidade dos dados. Processos de ETL mal projetados apresentam manutenção, alteração e atualização

dispendiosas. Conseqüentemente, é crucial fazer as escolhas adequadas em termos de tecnologia e ferramentas para usar no desenvolvimento e manutenção do processo de ETL.

São muitos os provedores de sistemas prontos de ETL. Os fornecedores de bancos de dados atualmente oferecem recursos de ETL que aprimoram e, ao mesmo tempo, competem com as ferramentas de ETL independentes. O SAS reconhece a importância da qualidade dos dados e oferece a primeira solução totalmente integrada da indústria, que une ETL e qualidade dos dados e transforma os dados em valiosos ativos estratégicos. Outros provedores de *software* de ETL incluem Microsoft, Oracle, IBM, Informatica, Embarcadero e Tibco. Para maiores informações sobre ETL, leia Eckerson (2003), Karacsony (2006) e Songini (2004).

Questões de revisão da Seção 2.5

1. Descreva a integração de dados.
2. Descreva as três etapas do processo de ETL.
3. Por que o processo de ETL é tão importante para os empreendimentos de data warehousing?

2.6

DESENVOLVIMENTO DE DATA WAREHOUSES

Um projeto de *data warehousing* é uma grande realização para qualquer empresa. Ele é mais complicado que um simples projeto de seleção e implementação de *mainframe*, pois abrange e influencia muitos departamentos e muitas interfaces de entrada e saída. O projeto pode também ser parte de uma estratégia comercial de CRM. Um *data warehouse* oferece alguns benefícios classificados como diretos e indiretos. Os benefícios diretos incluem:

- Permissão aos usuários finais que executem análises amplas de inúmeras formas.
- Uma visualização consolidada dos dados corporativos (isto é, uma versão única da verdade).
- Informações melhores e mais oportunas. Um *data warehouse* permite que o processamento de informações seja transferido dos caros sistemas operacionais para servidores de baixo custo; assim, muito mais solicitações de informação dos usuários finais podem ser processadas com mais rapidez.
- Desempenho aprimorado do sistema. Um *data warehouse* libera o processamento de produção porque alguns requisitos de relatórios do sistema operacional são movidos para o DSS.
- Simplificação do acesso aos dados.

Os benefícios indiretos resultam do uso feito dos benefícios diretos pelos usuários finais. No geral, esses benefícios melhoram o conhecimento dos negócios, apresentam vantagens competitivas, aprimoram o serviço e a satisfação do cliente, facilitam a tomada de decisões e ajudam a reformar os processos de negócios, sendo estas as mais fortes contribuições para a vantagem competitiva. (Para uma discussão sobre como criar uma vantagem competitiva através do *data warehousing*, consulte Parzinger e Frolick: 2001.) Leia em Watson et al. (2002) uma discussão detalhada sobre como as organizações podem obter excelentes níveis de recompensas. Dados os possíveis benefícios que um *data warehouse* pode oferecer e os investimentos consideráveis em tempo e dinheiro que tal projeto demanda, é essencial que uma organização estruture seu projeto para maximizar as chances de sucesso. Além disso, a organização obviamente deve levar os custos em consideração. Kelly (2001) descreveu uma abordagem ao retorno do investimento (ROI) que considera os benefícios nas categorias de *protetores* (i. e., dinheiro economizado pela melhoria das funções tradicionais de suporte à decisão), *coletores* (i. e., dinheiro economizado devido a coleta e disseminação automatizadas de informação) e *usuários* (i. e., dinheiro economizado ou ganho a partir de decisões tomadas com o uso do *data warehouse*). Os custos são relacionados a *hardware*, a *software*, a largura de banda, a desenvolvimento interno, a suporte interno, a treinamento e consultoria externa. O valor presente líquido é calculado sobre a vida útil esperada do *data warehou-*

se. Os benefícios são divididos aproximadamente em 20% para protetores, 30% para coletores e 50% para usuários; por isso, Kelly indica que os usuários devem estar envolvidos no processo de desenvolvimento, um fator de sucesso normalmente considerado crítico para os sistemas que implicam em mudança na organização.

O Caso de aplicação 2.3 mostra um exemplo de um *data warehouse* que foi desenvolvido e proporcionou uma grande vantagem competitiva para a Hokuriku Coca-Cola Bottling Company, no Japão. O sistema teve tanto sucesso que há planos de expandi-lo para as mais de um milhão de máquinas de venda de bebidas da Coca-Cola no país.

CASO DE APLICAÇÃO 2.3

As coisas melhoram com o *data warehouse* da Coca-Cola

Diante das pressões competitivas e da demanda dos consumidores, como uma empresa envasadora bem-sucedida garante os lucros de suas máquinas de venda? A resposta para a Hokuriku Coca-Cola Bottling Company (HCCBC) é um *software* de *data warehouse* e análise implementado pela Teradata Corp. A HCCBC construiu o sistema em resposta a um sistema de *data warehousing* desenvolvido por sua rival, Mikuni. Esse sistema coleta de cada máquina (considerada uma loja) não apenas dados históricos, mas também dados quase em tempo real, que podem ser transmitidos à sede via conexão sem fio. A fase inicial do projeto foi implantada em 2001. A abordagem ao *data warehouse* oferece informações detalhadas do produto, como data e hora de cada venda, quando um produto termina, se alguém deixou de receber troco, se a máquina está com defeito. Em cada caso, um *alerta* é disparado e a máquina de venda informa o centro de dados imediatamente por meio de um sistema de transmissão sem fio. (Note que a Coca-Cola nos Estados Unidos usou modems por mais de uma década para conectar as máquinas aos distribuidores.)

Em 2002, a HCCBC realizou um teste piloto e colocou todas as suas máquinas de vendas em Nagano em uma rede sem fio, para coletar dados quase em tempo real de cada ponto de venda (PDV).

Os resultados foram espantosos, pois fizeram uma previsão precisa da demanda e identificaram os problemas com rapidez. O total de vendas aumentou em 10% imediatamente. Além disso, devido à manutenção mais cuidadosa das máquinas, as horas extras e outros custos diminuíram em 46%. Cada vendedor pôde ainda fazer a manutenção de mais 42% de máquinas de venda.

O teste teve tanto êxito que o planejamento começou a ampliá-lo para compreender a empresa inteira (60.000 máquinas) usando um *data warehousing* ativo. Com o tempo, a solução de *data warehousing* irá ultrapassar as fronteiras corporativas e atingir todas as empresas envasadoras da Coca-Cola, para que as mais de um milhão de máquinas de vendas do Japão estejam em rede. Isso resultará em imensa economias de custo e maior receita.

Fontes: adaptado de K.D. Schwartz, “Decisions at the Touch of a Button”, *Teradata Magazine*, teradata.com/t/page/117774/index.html (acessado em abril 2006); K.D. Schwartz, “Decisions at the Touch of a Button”, *DSS Resources*, March 2004, pp. 28–31, dssresources.com/cases/coca-colajapan/index.html (acessado em abril de 2006); and Teradata Corp., *Coca-Cola Japan Puts the Fizz Back in Vending Machine Sales*, teradata.com/t/page/118866/index.html (acessado em abril de 2006).

Os aspectos essenciais para se ter um projeto bem-sucedido de *data warehousing* são definir claramente os objetivos, reunir suporte ao projeto junto à gerência e aos usuários finais, estabelecer cronogramas e orçamentos razoáveis e administrar as expectativas. Uma estratégia de *data warehousing* é um plano de ação para a introdução próspera desse tipo de sistema. A estratégia deve descrever aonde a empresa quer chegar, por que quer chegar lá e o que fará quando chegar lá. Ela precisa levar em consideração a visão, a estrutura e a cultura da organização. Veja em Matney (2003) os passos que podem ajudar a desenvolver uma estratégia de suporte flexível e eficiente. Quando forem estabelecidos o plano e o suporte de *data warehouse*, a organização deve examinar os fornecedores do sistema. (Consulte a Tabela 2.1 para ver uma lista com alguns fornecedores;

Tabela 2.1 Lista com alguns fornecedores de *data warehousing*

<i>Fornecedor</i>	<i>Oferta de produtos</i>
Computer Associates (cai.com)	Conjunto abrangente de ferramentas e produtos de <i>data warehouse</i> (DW)
DataMirror Corp. (datamirror.com)	Produtos de administração, gestão e desempenho de DW
Data Advantage Group, Inc. (dataadvantagegroup.com)	<i>Software</i> de metadados
Dell Computer Corp. (dell.com)	Servidores de DW
Embarcadero Technologies (embarcadero.com)	Produtos de administração, gestão e desempenho de DW
Business Objects (businessobjects.com)	<i>Software</i> de limpeza de dados
Harte-Hanks, Inc. (harte-hanks.com)	Produtos e serviços de gerenciamento de relacionamento com o cliente (CRM)
Hewlett-Packard Company (hp.com)	Servidores de DW
Hummingbird Ltd. (hummingbird.com)	Mecanismos de DW e exploração de <i>warehouses</i>
Hyperion Solutions Corp. (hyperion.com)	Conjunto abrangente de ferramentas, produtos e aplicações de DW
IBM (ibm.com)	Ferramentas, produtos e aplicações de DW
Informatica Corp. (informatica.com)	Produtos de administração, gestão e desempenho de DW
Microsoft Corp. (microsoft.com)	Ferramentas e produtos de DW
Oracle (inclui PeopleSoft e Siebel) (oracle.com)	Ferramentas, produtos e aplicações de DW, ERP e CRM
SAS Institute, Inc. (sas.com)	Ferramentas, produtos e aplicações de DW
Siemens (siemens.com)	Servidores de DW
Sybase, Inc. (sybase.com)	Conjunto abrangente de ferramentas e aplicações de DW
Teradata (teradata.com)	Ferramentas, produtos e aplicações de DW

veja também The Data Warehousing Institute [tdwi.org] e *DM Review* [dmreview.com].) Muitos fornecedores oferecem versões demo de seus produtos de *data warehousing* e BI.

Fornecedores de *data warehouse*

McCloskey (2002) citou seis diretrizes que devem ser consideradas ao se desenvolver uma lista de fornecedores: solidez financeira, vínculos com ERP, consultores qualificados, participação no mercado, experiência no setor e parcerias estabelecidas. É possível reunir mais dados em feiras comerciais e *websites* corporativos, e também solicitando informações sobre produtos específicos. Van den Hoven (1998) diferenciou três tipos de produtos de *data warehousing*. O primeiro tipo cuida de funções como localização, extração, transformação, limpeza, transporte e carga dos dados no *data warehouse*. O segundo tipo é uma ferramenta de gestão de dados: um mecanismo de banco de dados que armazena e gerencia o sistema e os metadados. O terceiro tipo é uma ferramenta de acesso aos dados que oferece acesso aos usuários finais para que analisem os dados no *warehouse*. Ela pode incluir recursos de gerador de consultas, visualização, EIS, OLAP e *data mining*.

Abordagens ao desenvolvimento do *data warehouse*

Muitas organizações necessitam criar os *data warehouses* usados para o suporte à decisão. Duas abordagens concorrentes são empregadas. A primeira delas é a de Bill Inmon, que é constantemente chamado de “pai do *data warehousing*”. Inmon apóia uma abordagem de desenvolvimento de cima para baixo que adapta as tradicionais ferramentas de banco de dados relacional às necessidades de desenvolvimento de um *data warehouse* para toda a empresa,

Tabela 2.2 Contrastes entre as abordagens de desenvolvimento EDW e de *data mart*

<i>Esforço</i>	<i>Abordagem de data mart</i>	<i>Abordagem EDW</i>
Escopo	Uma área temática	Várias áreas temáticas
Tempo de desenvolvimento	Meses	Anos
Custo de desenvolvimento	US\$ 10.000 a US\$ 100.000 ou mais	US\$ 1.000.000 ou mais
Dificuldade de desenvolvimento	Baixa a média	Alta
Pré-requisito de dados para compartilhamento	Comum (dentro da área de negócios)	Comum (em toda a empresa)
Fontes	Apenas alguns sistemas operacionais e externos	Muitos sistemas operacionais e externos
Tamanho	<i>Megabytes</i> a alguns <i>gigabytes</i>	<i>Gigabytes</i> a <i>petabytes</i>
Horizonte de tempo	Dados históricos e quase atuais	Dados históricos
Transformações de dados	Baixas a médias	Altas
Frequência de atualização	De hora em hora, diária, semanal	Semanal, mensal
<i>Tecnologia</i>		
<i>Hardware</i>	Estações de trabalho e servidores departamentais	Servidores empresariais e computadores <i>mainframe</i>
Sistema operacional	Windows e Linux	Unix, Z/OS, OS/390
Bancos de dados	Estações de trabalho ou servidores de banco de dados padrão	Servidores empresariais de banco de dados
<i>Uso</i>		
Número de usuários simultâneos	Dezenas	Centenas a milhares
Tipos de usuários	Analistas da área de negócios e gerentes	Analistas corporativos e altos executivos
Foco comercial	Otimizar as atividades dentro da área comercial	Otimização e tomada de decisões interfuncional

Fontes: adaptado de J. Van den Hoven, "Data Marts: Plan Big, Build Small", in *IS Management Handbook*, 8th ed., CRC Press, Boca Raton, FL, 2003; and T. Ariyachandra and H. Watson, "Which Data Warehouse Architecture Is Most Successful?" *Business Intelligence Journal*, Vol. 11, No. 1, First Quarter 2006, pp. 4–6.

conhecida também como *abordagem EDW*. A segunda é a de Ralph Kimball, que propõe uma abordagem de baixo para cima que emprega a modelagem dimensional, conhecida também como *abordagem de data mart*.

Conhecer de que forma estes dois modelos são semelhantes e diferentes ajuda a compreender os conceitos básicos de *data warehouse* (p. ex., veja Breslin: 2004). Mostraremos algumas vantagens e desvantagens das duas abordagens na Tabela 2.2. Descreveremos as abordagens detalhadamente a seguir.

O modelo de Inmon: a abordagem EDW

A abordagem de Inmon enfatiza o desenvolvimento de cima para baixo, empregando ferramentas e metodologias consagradas de desenvolvimento de banco de dados, como diagramas entidade-relacionamento (DER), e um ajuste da abordagem de desenvolvimento em espiral. A abordagem EDW não exclui a criação de *data marts*. O EDW é ideal nesta abordagem pois oferece uma visualização consistente e abrangente da empresa. Murtaza (1998) apresentou uma estrutura para desenvolver essa abordagem.

O modelo de Kimball: a abordagem de data mart

A estratégia de *data mart* de Kimball é uma abordagem “plano grande, construção pequena”. Um *data mart* é um *data warehouse* orientado por assunto ou departamento. É uma versão reduzida de um *data warehouse* que se concentra nas exigências de um departamento específico, como *marketing* ou vendas. Este modelo aplica a modelagem dimensional de dados, que começa com tabelas. Kimball defendeu uma metodologia de desenvolvimento que implica em uma abordagem de baixo para cima, o que, no caso dos *data warehouses*, significa construir um *data mart* por vez.

Qual modelo é melhor?

Não há uma estratégia única relativa ao *data warehousing*. A estratégia de uma empresa pode evoluir de um simples *data mart* até um complexo *data warehouse* em resposta às demandas dos usuários, às exigências de negócios da empresa e à sua maturidade na gestão de seus recursos de dados. Para muitas corporações, um *data mart* não raro é o primeiro passo conveniente para adquirir experiência na construção e gestão de um *data warehouse*, ao mesmo tempo em que apresenta

Tabela 2.3 Comparação das diferenças essenciais entre as características das abordagens de desenvolvimento de Inmon e Kimball

Característica	Inmon	Kimball
<i>Metodologia e arquitetura</i>		
Abordagem geral	De cima para baixo	De baixo para cima
Estrutura arquitetônica	<i>Data warehouse</i> para toda a empresa (atômico) “alimenta” os bancos de dados departamentais	Os <i>data marts</i> modelam um único processo de negócios, e chega-se à consistência da empresa por meio de um barramento de dados e do ajustar-se às dimensões
Complexidade do método	Bastante complexo	Bastante simples
Comparação com metodologias consagradas de desenvolvimento	Derivado da metodologia espiral	Processo de quatro passos; uma fase dos métodos de sistema de gerenciamento de banco de dados relacional (SGBDR)
Discussão do projeto físico	Bastante detalhada	Bastante leve
<i>Modelagem de dados</i>		
Orientação dos dados	Orientados por assunto ou por dados	Orientados por processos
Ferramentas	Tradicionais (diagramas entidade-relacionamento [ERD], diagramas de fluxo de dados [DFD])	Modelagem dimensional; uma fase da modelagem relacional
Acessibilidade ao usuário final	Baixa	Alta
<i>Filosofia</i>		
Público principal	Profissionais de TI	Usuários finais
Local na organização	Parte integral da fábrica de informações corporativa	Transformador e retentor dos dados operacionais
Objetivo	Proporcionar uma solução técnica sólida com base em métodos e tecnologias comprovados de banco de dados	Proporcionar uma solução que facilite aos usuários finais fazer consultas diretas aos dados e ainda obter tempos razoáveis de resposta

Fontes: adaptado de M. Breslin, “Data Warehousing Battle of the Giants: Comparing the Basics of Kimball and Inmon Models”, *Business Intelligence Journal*, Vol. 9, No. 1, Winter 2004, pp. 6–20; and T. Ariyachandra and H. Watson, “Which Data Warehouse Architecture Is Most Successful?” *Business Intelligence Journal*, Vol. 11, No. 1, First Quarter 2006.

aos usuários de negócios as vantagens de um acesso melhor aos dados. Ademais, é comum que um *data mart* indique o valor agregado do *data warehousing*. No final do processo, obter um EDW é o ideal (leia o Caso de aplicação 2.4). Entretanto, o desenvolvimento de *data marts* individuais com frequência gera muitas vantagens no caminho rumo ao desenvolvimento de um EDW, sobretudo se a organização não pode ou não quer investir em um projeto em grande escala. Os *data marts* também podem demonstrar a viabilidade e o sucesso em proporcionar benefícios, o que pode levar ao investimento em um EDW. A Tabela 2.3 resume as diferenças básicas nas características dos dois modelos.

CASO DE APLICAÇÃO 2.4

HP consolida centenas de *data marts* em um único EDW

Em dezembro de 2005, a Hewlett-Packard Co. (HP) planejou consolidar seus 762 *data marts* pelo mundo em um único EDW. A HP adotou esta abordagem para obter uma idéia melhor de seus negócios e também para determinar como servir melhor seus clientes. Mark Hurd, presidente e gerente-geral da HP, afirmou que “havia uma fome de dados analíticos” na empresa que, infelizmente, havia levado à criação de muitos *data marts*. Projetar e manter esses silos de dados era muito caro, sendo que eles não produziam a visão empresarial das informações internas e

de clientes desejada pela HP. Na metade de 2006, a HP começou a consolidar os dados dos *data marts* no novo *data warehouse*. Todos os *data marts* díspares terminarão sendo eliminados.

Fontes: adaptado de C. Martins, “HP to Consolidate Data Marts into Single Warehouse”, *Computerworld*, December 13, 2005; C. Martins, “HP to Consolidate Data Marts into Single Warehouse”, *InfoWorld*, December 13, 2005; e C. Martins, “HP to Consolidate Data Marts into One Warehouse”, *ITWorld Canada*, December 14, 2005.

Considerações adicionais sobre o desenvolvimento de *data warehouses*

Algumas organizações desejam terceirizar completamente seus empreendimentos de *data warehousing*. Elas simplesmente não querem lidar com aquisições de *hardware* e *software*, nem administrar seus sistemas de informação. Uma alternativa é usar *data warehouses* hospedados. Nesta situação, outra firma – idealmente, uma que tenha muita experiência e conhecimento – desenvolve e mantém o *data warehouse*. No entanto, esta abordagem traz preocupações com segurança e privacidade. Veja os *Insights* de tecnologia 2.5 para mais detalhes.

Estrutura de *data warehouse*: o esquema estrela

A Figura 2.1 mostra uma estrutura típica de *data warehouse*. Embora haja muitas variações da arquitetura, a mais importante delas é o esquema estrela. O projeto de *data warehouse* se baseia no conceito de *modelagem dimensional*. A **modelagem dimensional** é um sistema baseado em recuperação que suporta acessos com alto volume de consultas. O esquema estrela é o meio pelo qual a modelagem dimensional é implementada. Ele contém uma tabela de fatos central cercada por diversas tabelas de dimensão. A tabela de fatos contém uma grande quantidade de linhas que correspondem aos negócios ou fatos observados. Ela inclui os atributos necessários para executar análise de decisão, atributos descritivos usados para geração de relatórios sobre as consultas e chaves externas para vinculação a tabelas de dimensão. Os atributos da análise de decisão consistem em medidas de desempenho, métricas operacionais, medidas agregadas e todas as outras métricas necessárias para analisar o desempenho da organização. Em outras palavras, a tabela de fatos trata principalmente de *o que* o *data warehouse* suporta na análise de decisão.

Que tal um *data warehouse* hospedado?

Um *data warehouse* hospedado tem quase as mesmas (se não mais) funcionalidades de um *data warehouse* local, mas não consome recursos computacionais nas dependências do cliente. Ele oferece os benefícios da BI, menos o custo com atualizações de computadores ou redes, licenças de *software*, e desenvolvimento, manutenção e suporte internos.

Um *data warehouse* hospedado oferece os seguintes benefícios:

- Exige um investimento mínimo em infra-estrutura.
- Libera capacidade dos sistemas internos.
- Libera o fluxo de caixa.
- Torna acessíveis soluções eficientes.
- Possibilita soluções eficientes que proporcionam crescimento.
- Oferece melhor qualidade de equipamentos e *softwares*.
- Propicia conexões mais rápidas.

- Habilita os usuários a acessarem os dados de locais remotos.
- Permite que a empresa se concentre nos negócios essenciais.
- Atende às necessidades de armazenamento de grandes volumes de dados.

Apesar de seus benefícios, um *data warehouse* hospedado não é obrigatoriamente uma boa opção para todas as empresas. Grandes empresas com receita superior a US\$ 500 milhões podem perder dinheiro se já possuem infra-estrutura e equipe de TI internas subutilizadas. Além do mais, as empresas que vêm a mudança de paradigma de aplicações terceirizadas como uma perda de controle de seus dados, provavelmente não usarão um provedor de serviços de *business intelligence* (BISP). Por fim, o argumento mais significativo e comum contra a implementação de um *data warehouse* é que pode ser imprudente terceirizar aplicações confidenciais por motivos de segurança e privacidade.

Fontes: adaptado parcialmente de M.Thornton and M. Lampa, "Hosted Data Warehouse", *Journal of Data Warehousing*, Vol. 7, No. 2, 2002, pp. 27–34; and M.Thornton, "What About Security? The Most Common, but Unwarranted, Objection to Hosted Data Warehouses", *DM Review*, Vol. 12, No. 3, March 18, 2002, pp. 30–43.

Ao redor das tabelas de fatos centrais (e vinculadas pelas chaves externas) estão as *tabelas de dimensão*. As **tabelas de dimensão** contém informações de classificação e agregação sobre as linhas de fatos centrais. Elas incluem atributos que descrevem os dados contidos na tabela de fatos e tratam de *como* os dados serão analisados. Estas tabelas têm uma relação de um para muitos com as linhas na tabela de fatos central. Alguns exemplos de dimensões que suportariam uma tabela de fatos de produto são localização, tempo e tamanho. O esquema estrela oferece tempo extremamente rápido de resposta a consultas, simplicidade e facilidade de manutenção para estruturas de bancos de dados somente leitura. Segundo Raden (2003), estabelecer um esquema estrela para atualizações em tempo real pode ser uma abordagem objetiva, contanto que algumas regras sejam seguidas. Mostramos um exemplo de esquema estrela na Figura 2.9.

O **grão** (conhecido também como granularidade) de um *data warehouse* define o nível mais alto de detalhes suportado. Ele indica se o *data warehouse* é altamente resumido ou se também inclui dados detalhados sobre transações. Se o grão for muito alto, o *warehouse* pode não admitir solicitações para fazer *drill down* nos dados. A análise **drill down** é o processo de examinar mais além de um valor resumido para investigar cada uma das transações detalhadas que compõem o resumo. Um baixo nível de granularidade resultará em mais dados armazenados no *warehouse*. Quantidades maiores de detalhes podem afetar o desempenho das consultas deixando o tempo de resposta mais longo. Assim, durante a definição do escopo de um projeto de *data warehouse*, é importante identificar o nível justo de granularidade que será necessário. Consulte Tennant (2002) para ler uma discussão sobre questões de granularidade nos metadados.

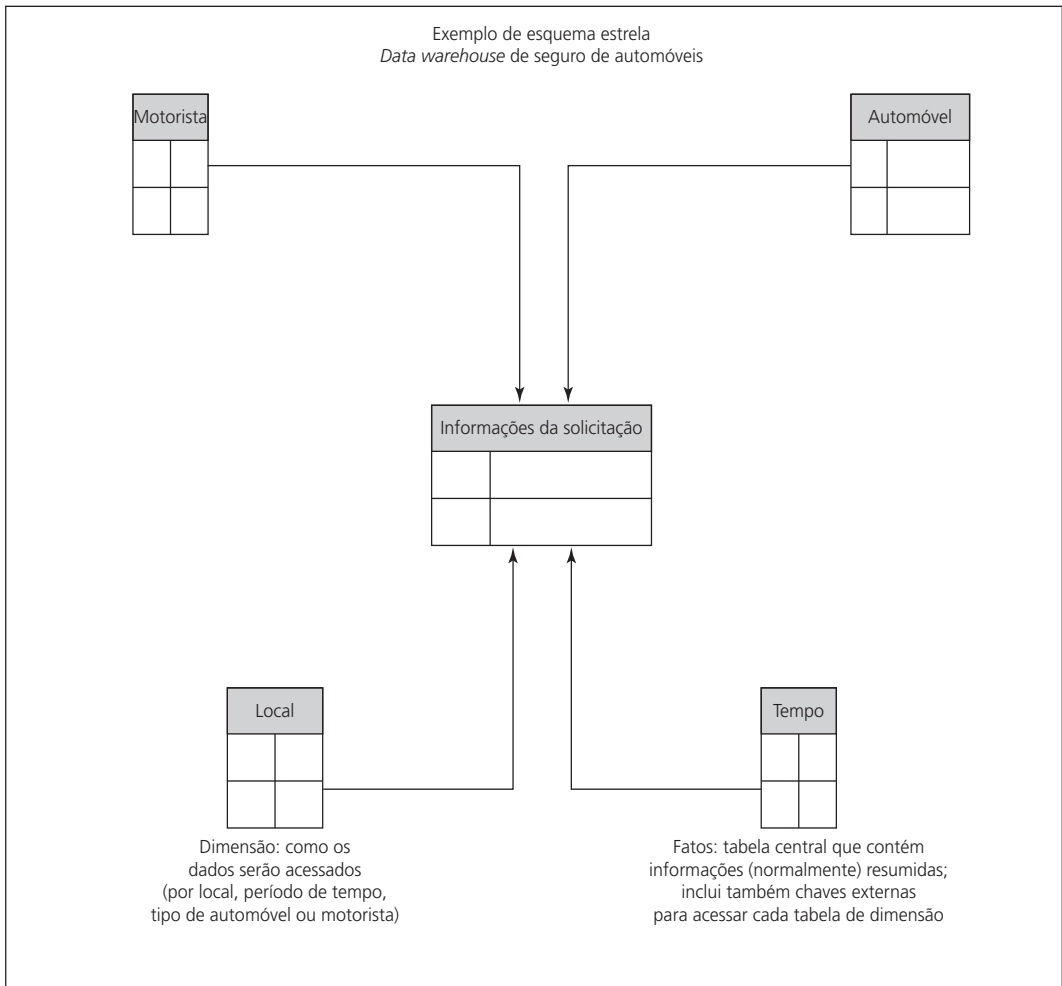


Figura 2.9 Esquema estrela.

Questões sobre a implementação de data warehousing

Implementar um *data warehouse* normalmente é um esforço pesado que deve ser planejado e executado de acordo com métodos estabelecidos. No entanto, o ciclo de vida do projeto tem muitas facetas e uma única pessoa não pode ser especialista em todas as áreas. Discutiremos aqui idéias e questões específicas que se relacionam ao *data warehousing*. Inmon (2006) propôs um conjunto de ações que um programador de sistemas de *data warehouse* pode usar para ajustar um *warehouse*.

Solomon (2005) ofereceu algumas diretrizes relativas a perguntas críticas que devem ser feitas, alguns riscos que devem ser ponderados e alguns processos que podem ser seguidos para se garantir uma implementação bem-sucedida do *data warehouse*. Ele compilou uma lista de 11 tarefas principais a serem executadas paralelamente:

1. Estabelecimento de acordos de nível de serviço e solicitações para atualização de dados
2. Identificação de fontes de dados e suas políticas de controle
3. Planejamento da qualidade dos dados
4. Projeto de modelo de dados
5. Seleção de ferramentas de ETL

6. Seleção de plataforma e *software* de banco de dados relacional
7. Transporte de dados
8. Conversão de dados
9. Processo de reconciliação
10. Limpeza e planejamento de arquivo
11. Suporte ao usuário final

Seguir essas diretrizes deve aumentar as chances de uma organização alcançar o sucesso. Dado o tamanho e escopo de uma iniciativa de *data warehouse* no nível da empresa, a não-antecipação dessas questões aumenta em muito os riscos de fracasso.

Hwang e Xu (2005) realizaram uma grande pesquisa sobre as questões que resultam no sucesso do *data warehouse*. Os resultados estabeleceram que o sucesso nesta área é um conceito multifacetado, e os autores propuseram que um *data warehouse* seja construído mantendo-se em mente o objetivo de aumentar a produtividade do usuário. As vantagens significativas de agir dessa forma incluem pronta recuperação das informações e melhor qualidade da informação. Os resultados da pesquisa indicaram também que o êxito da implementação depende de fatores de diferentes dimensões.

As pessoas desejam saber se suas iniciativas de BI e *data warehousing* são bem-sucedidas em comparação às de outras empresas, portanto, Ariyachandra e Watson (2006a) propuseram alguns referenciais para o sucesso dessas iniciativas. Já Watson et al. (1999) pesquisaram os fracassos de *data warehousing* e seus resultados mostraram que as pessoas definem “fracasso” de diversas formas, o que foi confirmado por Ariyachandra e Watson (2006a). O The Data Warehousing Institute (**tdwi.org**) desenvolveu um modelo de maturidade de *data warehousing* que uma empresa pode aplicar a fim de marcar referenciais de sua evolução. Ele oferece uma forma rápida de medir onde a iniciativa de *data warehousing* da empresa encontra-se agora e aonde precisa ir a seguir. O modelo de maturidade consiste em seis estágios: pré-natal, infante, criança, adolescente, adulto e sábio. O valor comercial cresce conforme o *warehouse* avança por cada estágio. Os estágios são identificados por diversas características, incluindo escopo, estrutura analítica, percepções executivas, tipos de análises, liderança, recursos financeiros, plataforma tecnológica, gestão de mudanças e administração. Para os detalhes completos, veja Eckerson (2004).

Weir (2002) descreveu algumas das melhores práticas para implementar um *data warehouse*, que incluem as seguintes diretrizes:

- O projeto deve ser adequado à estratégia corporativa e aos objetivos de negócios.
- Deve haver adesão completa ao projeto por parte dos executivos, gerentes e usuários.
- É importante gerenciar as expectativas do usuário sobre o projeto concluído.
- O *data warehouse* deve ser construído progressivamente.
- Deve ser incorporada a adaptabilidade.
- Tanto profissionais de TI quanto de negócios devem gerir o projeto.
- Deve ser desenvolvida uma relação empresa/fornecedor.
- Devem ser carregados somente os dados que foram limpos ou que sejam de uma qualidade conhecida pela organização.
- Não se deve negligenciar as exigências de treinamento.
- É preciso estar politicamente consciente.

Os projetos de *data warehouse* envolvem muitos riscos. A maioria destes também é encontrada em outros projetos de TI, mas os riscos com *data warehousing* são mais sérios por que são projetos caros e em grande escala. Cada risco deve ser avaliado logo no seu início. Adelman e Moss (2001) descreveram alguns desses riscos, que incluem:

- Ausência de missão ou objetivo
- Qualidade dos dados-fonte desconhecida
- Habilidades não estão a postos

- Orçamento inadequado
- Falta de *software* de suporte
- Dados-fonte incompreendidos
- Patrocinador fraco
- Usuários não têm conhecimentos de informática
- Problemas políticos ou disputas territoriais
- Usuários não têm expectativas realistas
- Riscos arquitetônicos e de projeto
- Aumento do escopo e exigências que mudam
- Fornecedores fora de controle
- Múltiplas plataformas
- Pessoas essenciais deixam o projeto
- Perda de patrocinador
- Demasiada Tecnologia nova
- Necessidade de alterar um sistema operacional
- Ambiente distribuído por vários locais
- Cultura geográfica e idiomática da equipe

Os praticantes revelaram uma profusão de erros cometidos no desenvolvimento de *data warehouses*. Watson et al. (1999) discutiram como esses erros poderiam levar a fracassos no sistema (leia também Barquin et al.: 1997). Turban et al. (2006) listaram fatores de fracasso que incluem questões culturais ignoradas, arquitetura inadequada, objetivos de negócio incertos, informações ausentes, expectativas não-realistas, baixos níveis de resumo de dados e baixa qualidade dos dados.

Ao desenvolver um *data warehouse* bem-sucedido, é importante considerar atentamente estas questões:

- *Começar com a cadeia de patrocinadores errada.* É preciso ter um patrocinador executivo que tenha influência sobre os recursos necessários, a fim de investir no *data warehouse* e apoiá-lo. Também é necessário um *condutor do projeto*, alguém que tenha conquistado o respeito dos outros executivos, possua um ceticismo saudável sobre tecnologia e seja decisivo porém flexível. E você precisa de um gerente de IS/TI para liderar o projeto (o “você” do projeto).
- *Criar expectativas que não podem ser atingidas e frustrar os executivos na hora da verdade.* Há duas fases em cada projeto de *data warehousing*. A fase 1 é a fase da venda, em que você comercializa o projeto internamente, vendendo as vantagens para aqueles que têm acesso aos recursos necessários. A fase 2 é a luta para atingir as expectativas descritas na fase 1. Desde um mero dólar até 7 milhões, espera-se que você cumpra com o prometido.
- *Assumir um comportamento politicamente inocente.* Não afirme simplesmente que um *data warehouse* ajudará os gerentes a tomarem decisões melhores: isso pode implicar que você ache que eles, até agora, tomavam decisões ruins. Venda a idéia de que eles poderão obter as informações necessárias para auxiliar na tomada de decisões.
- *Carregar informações no warehouse só porque ele está disponível.* Não deixe o *data warehouse* se tornar uma lixeira de dados. Isso tornaria o uso do sistema mais lento sem necessidade. Há uma tendência voltada à computação e à análise em tempo real. Os *data warehouses* devem ser fechados e, automaticamente carregarem os dados de forma oportuna.
- *Acreditar que o projeto de banco de dados de data warehousing é o mesmo do projeto de banco de dados transacional.* Normalmente, não é. O objetivo do *data warehousing* é acessar registros agregados, e não um registro individual ou poucos registros, como nos sistemas de processamento de transações. O conteúdo também difere, como fica evidente pela forma como os dados são organizados. O SGBD tende a ser não-redundante, normalizado e relacional, enquanto os *data warehouses* são redundantes, não-normalizados e multidimensionais.

- *Escolher um gerente de data warehouse que seja orientado pela tecnologia, não pelo usuário.* Uma chave para o sucesso em *data warehouses* é compreender que os usuários devem obter o que precisam, e não se deve adotar tecnologia avançada apenas em nome da tecnologia.
- *Concentrar-se nos tradicionais dados internos orientados por registros, e ignorar o valor de dados externos e de textos, imagens e, talvez, som e vídeo.* Os dados vêm em muitos formatos e devem ser disponibilizados às pessoas certas, na hora certa e no formato certo. Eles devem ser catalogados adequadamente.
- *Oferecer dados com definições confusas e conflitantes.* A limpeza dos dados é um aspecto vital do *data warehousing*. Ela inclui o ajuste dos conflitos em definições e formatos dos dados em toda a empresa. Isto pode ser difícil politicamente pois envolve mudanças, normalmente no nível executivo.
- *Acreditar em promessas de desempenho, capacidade e escalabilidade.* Os *data warehouses* geralmente exigem mais capacidade e velocidade do que estava previsto no orçamento original. Faça planos futuros de ampliação.
- *Acreditar que seus problemas terminarão quando o data warehouse estiver funcionando.* Os projetos de DSS/BI tendem a evoluir continuamente. Cada implementação é uma iteração do processo de prototipagem. Sempre haverá necessidade de adicionar mais e mais conjuntos de dados ao *warehouse*, bem como outras ferramentas de análise para grupos novos ou existentes de tomadores de decisão. O planejamento deve incluir uma grande energia e orçamentos anuais, pois o sucesso rende sucesso. O processo de *data warehousing* é contínuo.
- *Concentrar-se em data mining e relatórios periódicos ad hoc em vez de alertas.* O progresso natural das informações em um *data warehouse* é: 1) *extrair* os dados dos sistemas legados, limpá-los e usá-los para alimentar o *warehouse*; 2) *oferecer suporte* aos relatórios *ad hoc* até você aprender o que as pessoas querem; 3) *converter* os relatórios *ad hoc* em relatórios agendados regularmente.

Este processo de aprender o que as pessoas querem para depois poder dá-lo a elas parece natural, mas não é ideal e nem prático. Os gerentes são pessoas ocupadas e necessitam de tempo para ler os relatórios. *Sistemas de alerta* são melhores que sistemas de relatórios periódicos e podem tornar essencial o trabalho com o *data warehouse*. Esses sistemas monitoram os dados que fluem ao *warehouse* e dão informações a todas as pessoas-chave que precisam saber assim que um evento crítico ocorre.

Sammon e Finnegan (2000) revelaram o resultado de um estudo com quatro usuários maduros de tecnologia de *data warehousing*. Suas práticas foram resumidas em um esboço de 10 requisitos organizacionais para a aplicação do sistema. Os autores acreditam que as organizações poderiam utilizar essa representação para avaliar internamente as chances de sucesso de um projeto de *data warehousing*, e também para reconhecer as partes que precisam de atenção antes do início da implementação. Segue um resumo de seus pré-requisitos:

- Uma iniciativa de *data warehousing* orientada aos negócios.
- Patrocínio e compromisso executivos.
- Compromisso de recursos financeiros baseado em expectativas gerenciadas de forma realista.
- Uma equipe de projeto.
- Atenção à qualidade dos dados-fonte.
- Um modelo flexível de dados corporativos.
- Liderança de dados.
- Um plano de longo prazo para métodos/ferramentas de extração automatizada de dados.
- Conhecimento sobre a compatibilidade do *data warehouse* com os sistemas existentes.
- Prova de conceito de *hardware/software*.

Wixom e Watson (2001) definiram um modelo de pesquisa para o sucesso com *data warehouses* que identificou sete fatores importantes da implementação, os quais podem ser categorizados em três critérios (questões organizacionais, questões de projeto e questões técnicas):

1. Apoio administrativo
2. Defensores
3. Recursos
4. Participação dos usuários
5. Habilidades da equipe
6. Sistemas-fonte
7. Tecnologia de desenvolvimento

Em muitas organizações, um *data warehouse* terá sucesso apenas se houver um forte apoio da alta gerência ao seu desenvolvimento e se houver um defensor do projeto. Embora provavelmente aconteça em qualquer projeto de TI, isso é importante sobretudo para esse tipo de sistema. A implementação triunfante de um *data warehouse* resulta no estabelecimento de uma estrutura arquitetônica que pode proporcionar a análise de decisão por toda uma organização. Em alguns casos, ela oferece uma abrangente gestão da cadeia de fornecimento (SCM) garantindo acesso aos clientes e fornecedores da organização. A implementação de *data warehouses* baseados na Web (*webhousing*) facilitou o acesso a vastas quantidades de dados, mas é difícil determinar os benefícios concretos associados a esse sistema. *Benefícios concretos* são definidos como os benefícios a uma organização que podem ser expressos em termos financeiros. Muitas organizações têm recursos limitados de TI e devem priorizar seus projetos. O apoio administrativo e um forte defensor do projeto podem ajudar a garantir que o *data warehouse* receba os recursos necessários para sua boa implementação que podem representar um custo significativo, em alguns casos exigindo processadores avançados e grandes aumentos nos dispositivos de armazenamento de acesso direto (DASD). *Data warehouses* baseados na Web podem ainda ter requisitos de segurança para que apenas usuários autorizados acessem os dados.

A participação do usuário no desenvolvimento da modelagem de dados e do acesso é um fator crítico para o sucesso no desenvolvimento do *data warehouse*. Durante a modelagem dos dados, é preciso conhecimento especializado para determinar quais dados são necessários, definir regras de negócios associadas a eles e decidir que agregações e outros cálculos podem ser necessários. A modelagem do acesso é necessária para determinar como os dados serão recuperados de um *warehouse* e auxilia na sua definição física, ajudando a determinar quais dados precisam de indexação. Ela pode igualmente indicar se são precisos *data marts* dependentes para simplificar a recuperação das informações. As habilidades da equipe necessárias para desenvolver e implementar um *data warehouse* incluem conhecimento profundo sobre a tecnologia de banco de dados e sobre as ferramentas de desenvolvimento a serem usadas. Os sistemas-fonte e a tecnologia de desenvolvimento, como dissemos anteriormente, referem-se aos muitos dados de entrada e processos usados para carregar e manter um *data warehouse*.

***Data warehouses* gigantescos e escalabilidade**

Além da flexibilidade, um *data warehouse* precisa permitir escalabilidade. As principais questões ligadas à escalabilidade são a quantidade de dados no *warehouse*, a rapidez com que se espera que ele cresça, o número de usuários simultâneos e a complexidade das consultas dos usuários. Um *data warehouse* deve ter escalabilidade tanto horizontal quanto vertical. O *warehouse* crescerá em função do aumento dos dados e da necessidade de expandi-lo para dar suporte a novas funcionalidades de negócios. O crescimento dos dados pode ser resultado da adição de dados do ciclo atual (p. ex., os resultados deste mês) e/ou de dados históricos.

Hicks (2001) descreveu grandes bancos de dados e *data warehouses*. O Wal-Mart está sempre aumentando o tamanho de seu gigantesco *data warehouse*. A corporação usa um *warehouse* com

centenas de *terabytes* de dados para estudar as tendências de vendas e acompanhar o estoque e outras tarefas. O Departamento de Defesa dos Estados Unidos está usando um *data warehouse* e repositório de 5 *petabytes* para armazenar os registros médicos de 9 milhões de militares. Devido ao armazenamento exigido para arquivar suas filmagens, a CNN também possui um *data warehouse* que atingiu a marca dos *petabytes*.

Considerando que o tamanho dos *data warehouses* cresce em ritmo exponencial, a *escalabilidade* é uma questão importante. Boa escalabilidade significa que as consultas e outras funções de acesso aos dados (idealmente) terão um crescimento diretamente proporcional ao tamanho do *warehouse*. Consulte Rosenberg (2006) para ler abordagens à melhoria do desempenho das consultas. Na prática, foram desenvolvidos métodos especializados para criar *data warehouses* escalonáveis. É difícil obter escalabilidade ao se gerenciarem centenas de *terabytes* ou mais. *Tera-bytes* de dados têm uma inércia considerável, ocupam muito espaço físico e exigem computadores potentes. Algumas firmas usam processamento paralelo, outras usam esquemas inteligentes de indexação e pesquisa para gerenciar seus dados. Outras ainda espalham seus dados por diferentes *data stores* físicos. Conforme mais *warehouses* se aproximam dos *petabytes*, são desenvolvidas soluções cada vez melhores para a obtenção da escalabilidade.

Hall (2002) também tratou das questões de escalabilidade. A AT&T é uma líder do setor na implementação e uso de *data warehouses* gigantescos. Com seu *warehouse* de 26 *terabytes*, a AT&T Labs pode detectar o uso fraudulento de cartões telefônicos e investigar ligações relacionadas a seqüestros e outros crimes. Ela também consegue computar milhões de votos populares dos telespectadores que elegerão o vencedor do American Idol.

Para uma amostra de implementações bem-sucedidas de *data warehousing*, consulte Edwards (2003). Jukic e Lang (2004) examinaram as tendências e questões específicas relacionadas ao uso de recursos estrangeiros no desenvolvimento e suporte de aplicações de BI e *data warehousing*. Davison (2003) indicou que a terceirização no exterior relativa a TI crescia de 20 a 25% ao ano. Quando considerar projetos de *data warehousing* com recursos estrangeiros, considere muito bem as questões de cultura e segurança (consulte Jukic E Lang: 2004 para maiores detalhes).

Questões de revisão da Seção 2.6

1. Liste as vantagens dos *data warehouses*.
2. Liste alguns critérios para selecionar um fornecedor de *data warehouse* e descreva por que eles são importantes.
3. Uma abordagem de desenvolvimento de *data warehouses* de baixo para cima usa um modelo de dados corporativos?
4. Descreva as principais semelhanças e diferenças entre as abordagens de desenvolvimento de *data warehouses* de Inmon e Kimball.
5. Liste os diferentes tipos de arquiteturas de *data warehouse*.

2.7

DATA WAREHOUSING EM TEMPO REAL

As ferramentas de *data warehousing* e BI tradicionalmente se concentram em auxiliar os gerentes na tomada de decisões estratégicas e táticas. Maiores volumes de dados e velocidades mais rápidas de atualização estão alterando a sua função nas empresas modernas. Para muitas empresas, tomar decisões velozes e consistentes em toda a corporação exige mais do que um *data warehouse* ou *data mart* tradicional, esses não são essenciais aos negócios. Os dados são comumente atualizados a cada semana, e isso não permite uma resposta às transações quase em tempo real.

Mais dados, que vêm mais rapidamente e precisam ser convertidos em decisões, significam que as organizações se deparam com a necessidade de ter *data warehousing* em tempo real. Isso

porque o suporte à decisão tornou-se operante, a BI integrado exige análise de ciclo fechado, e os ODS do passado não oferecem suporte às exigências atuais.

Em 2003, com o advento do *data warehousing* em tempo real, essas tecnologias passaram a ser usadas nas decisões operacionais. O **data warehousing em tempo real (RDW)**, conhecido também como **data warehousing ativo (ADW)**, é o processo de carregar e fornecer dados por meio do *data warehouse* conforme eles se tornam disponíveis. Apenas recentemente ele evoluiu a partir do conceito de EDW. As características ativas de um RDW/ADW complementam e ampliam as funções do sistema tradicional relativas à tomada de decisões táticas. Por toda a organização, as pessoas que interagem diretamente com clientes e fornecedores terão autonomia para tomar decisões baseadas em informações ao alcance das mãos. Um uso ainda mais intenso dos resultados será possível quando o ADW proporcionar as informações diretamente aos clientes e fornecedores. O alcance e o impacto do acesso às informações na tomada de decisões podem afetar de forma positiva quase todos os aspectos do serviço ao cliente, SCM, logística e outros. O comércio eletrônico tornou-se um grande catalisador da demanda por *data warehousing* ativo (veja Armstrong: 2000). Por exemplo, a loja virtual Overstock.com, Inc. (**overstock.com**) conectou os usuários dos dados a um *data warehouse* em tempo real. No Egg plc, o maior banco exclusivamente *online* do mundo, um *data warehouse* de clientes é atualizado quase em tempo real. Consulte o Caso de aplicação 2.5.

CASO DE APLICAÇÃO 2.5

O Egg plc arrasa a concorrência quase em tempo real

O Egg plc (**egg.com**) é o maior banco *online* do mundo. Ele oferece serviços de banco, seguros, investimentos e hipotecas para mais de 3,6 milhões de clientes através de seu *site* na Internet. Em 1998, o Egg escolheu a Sun Microsystems para criar uma infra-estrutura confiável, escalonável e segura para dar suporte aos seus mais de 2,5 milhões de transações diárias. Em 2001, o sistema recebeu um *upgrade* a fim de eliminar problemas de latência. Esse novo *data warehouse* de clientes (CDW) usou *softwares* da Sun, Oracle e SAS. O *warehouse* inicial tinha cerca de 10 *terabytes* de dados e usava um servidor de 16 CPUs. O sistema oferece acesso aos dados quase em tempo real. Ele proporciona também serviços de *data warehouse* e *data mining* aos usuários internos, além de um conjunto solicitado

de dados dos clientes aos próprios clientes. Centenas de campanhas de vendas e *marketing* são elaboradas com base nos dados quase em tempo real (em alguns minutos). E, melhor ainda, o sistema permite que se tomem mais rapidamente decisões sobre clientes e classes de clientes específicos.

Fontes: compilado de “Egg’s Customer Data Warehouse Hits the Mark”, *DM Review*, Vol. 15, No. 10, October 2005, pp. 24–28; Sun Microsystems, *Egg Banks on Sun to Hit the Mark with Customers*, September 19, 2005, sun.com/smi/Press/sunflash/2005-09/sunflash.20050919.1.xml (acessado em abril de 2006); and ZD Net UK, *Sun Case Study: Egg’s Customer Data Warehouse*, whitepapers.zdnet.co.uk/0,39025945,60159401p-39000449q,00.htm (acessado em abril de 2006).

Conforme evoluem as necessidades de negócios, evoluem também as exigências do *data warehouse*. Neste nível básico, um *data warehouse* simplesmente relata o que ocorreu. No nível seguinte, há alguma análise. Com a evolução do sistema, ele proporciona recursos de previsão, que levam ao próximo nível de operacionalização. No seu nível mais alto de evolução, o ADW é capaz de fazer os eventos acontecerem (p. ex., atividades como criação de campanhas de vendas e *marketing*, ou identificação e exploração de oportunidades). Veja na Figura 2.10 uma descrição gráfica deste processo evolutivo.

A Teradata Corp. oferece os requisitos básicos para o suporte a um EDW. Ela fornece também as novas características de *data warehousing* ativo exigidas para proporcionar dados novos, dis-

Inteligência empresarial ativa: novidades

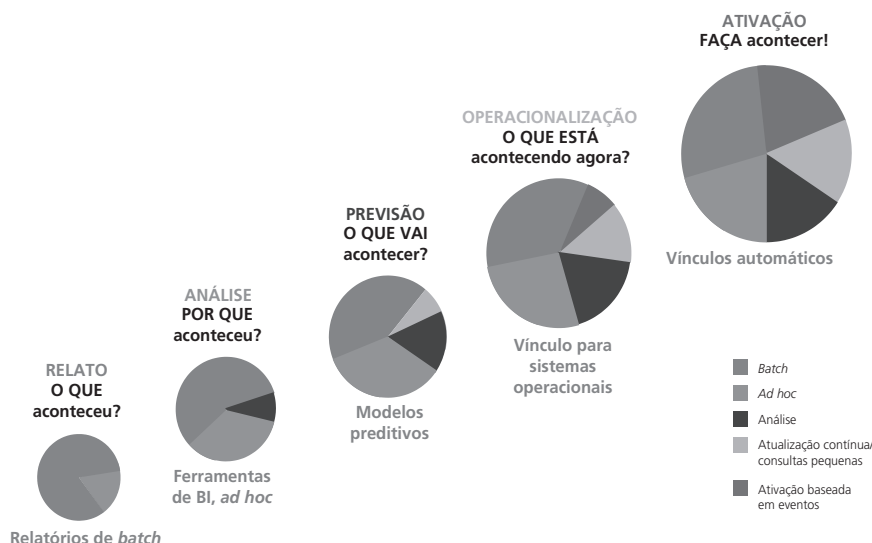


Figura 2.10 A evolução das decisões corporativas.

Fonte: cortesia de Teradata Corporation, uma divisão da NCR Corp. Usado mediante permissão.

poníveis e com bom desempenho, além de permitir a gestão das decisões corporativas (consulte a Figura 2.11 para um exemplo).

Um ADW oferece um repositório integrado de informações para promover o suporte à decisão estratégica e tática dentro de uma organização. Em vez da extração de dados operacionais em lotes noturnos de um sistema OLTP para um ODS, o *data warehousing* em tempo real defende que os dados sejam reunidos dos sistemas OLTP, conforme e quando os eventos acontecem, e movidos de uma só vez para o *data warehouse*. Isso permite a atualização instantânea do *warehouse* e a eliminação de um ODS. Neste ponto, podem ser realizadas consultas táticas e estratégicas no RDW para o uso de dados imediatos e históricos.

Segundo Basu (2003), a diferença mais marcante entre um *data warehouse* tradicional e um RDW é a mudança de paradigma de aquisição de dados. Entre os casos de negócios e exigências empresariais que levaram a essa necessidade de dados em tempo real estão os seguintes:

- Uma empresa nem sempre pode se dispor a esperar um dia inteiro para carregar seus dados operacionais no *warehouse* e fazer a análise.
- Até agora, os *data warehouses* captaram instantâneos dos estados fixos de uma organização, e não dados progressivos em tempo real exibindo cada mudança de estado e padrões quase análogos com o tempo.
- Com uma arquitetura em estrela tradicional, é difícil conservar os metadados em sintonia. Também custa caro desenvolver, manter e segurar muitos sistemas, ao contrário de um único *data warehouse* grande, onde os dados são centralizados para as ferramentas de BI/BA.

Data warehouse ativo da Teradata?

Uma extensão "ativa" do data warehouse empresarial

Carga ativa

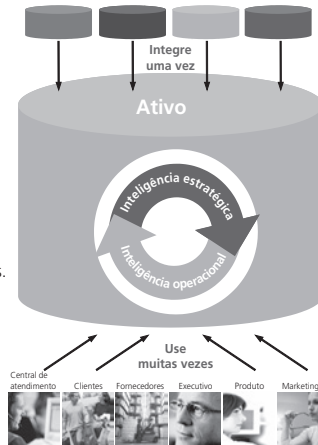
Aquisição de dados do dia; transmissão de dados, de mini-lotes a pequenos fluxos quase em tempo real (NRT), medida em minutos ou segundos.

Acesso ativo

Decisões ou serviços operacionais de linha de frente suportados pelo acesso NRT; acordos de nível de serviço de 5 segundos ou menos.

Eventos ativos

Monitoramento proativo da atividade comercial inicia ações inteligentes com base em regras e contexto; para sistemas ou usuários que apoiem um processo de negócios operante.



Gerenciamento ativo da carga de trabalho

Gerencie de forma dinâmica os recursos do sistema para obter utilização de recursos e desempenho ideais, apoiando um ambiente com cargas de trabalho mistas.

Integração empresarial ativa

Integração com a arquitetura empresarial para fornecer serviços inteligentes de decisões.

Disponibilidade ativa

Continuidade dos negócios para arcar com as exigências comerciais (até 24 x 7 x 365).

Figura 2.11 O EDW ativo da Teradata.

Fonte: cortesia de Teradata Corporation, uma divisão da NCR Corp. Usado mediante permissão.

- Em casos de cargas de lotes noturnos volumosos, pode ser muito alta a energia necessária para processamento e configuração de ETL em grandes carregamentos noturnos do *data warehouse*. Além disso, os processos podem demorar demais. Uma integração de aplicações corporativas (EAI) com coleta de dados em tempo real consegue reduzir ou eliminar os processos de lotes noturnos.

Apesar dos benefícios de um RDW, o seu desenvolvimento pode criar uma boa cota de problemas, relacionados a arquitetura; modelagem de dados; projeto, armazenamento e escalabilidade do banco de dados físico; e manutenibilidade. Ademais, dependendo de quando exatamente os dados são acessados, e contando até mesmo os microssegundos, podem ser extraídas e criadas diferentes versões da verdade, situação que confundirá os membros da equipe. Para maiores detalhes, leia Basu (2003) e Terr (2004).

As soluções em tempo real apresentam desafios notáveis às atividades de BI. Embora não seja o ideal para todas as soluções, o *data warehousing* em tempo real pode ter sucesso se a organização desenvolver uma metodologia segura para lidar com os riscos do projeto, incorporar um planejamento adequado e se concentrar em atividades de garantia da qualidade. Compreender os desafios comuns e aplicar melhores práticas podem reduzir os níveis de problemas que, não raro, fazem parte da implementação de sistemas complexos de *data warehousing* que incorporam métodos de BI/BA. Burdett e Singh (2004) e Wilk (2003) discutem os detalhes e implementações reais; veja também Akbay (2006) e Ericson (2006).

Confira os *Insights* de tecnologia 2.7 para conhecer alguns detalhes da evolução do conceito de tempo real. A aplicação de *dashboard* para gerenciamento de vôos na Continental Airlines (consulte a vinheta de abertura) ilustra o poder da BI em tempo real em acessar o *data warehouse* para uso em situações de interação direta com o cliente. A equipe de operações usa o sistema em tempo real para identificar problemas na rede de vôos da empresa. Outro exemplo: a UPS

As realidades em tempo real do data warehousing ativo

Em 2003, o papel do data warehousing na prática crescia rapidamente. Os sistemas em tempo real, ainda que fossem uma novidade, eram o assunto do momento, junto com as grandes complicações de oferta de dados e informações de forma instantânea a quem precisasse. Muitos especialistas, incluindo Peter Coffee, editor de tecnologia da revista *eWeek*, acreditam que os sistemas em tempo real devem fomentar um processo de tomada de decisões em tempo real. Stephen Brobst, CTO da Teradata, divisão da NCR, indicou que o *data warehousing* ativo é um processo de evolução da forma como uma empresa usa os dados. *Ativo* quer dizer que o data warehouse também é usado como uma ferramenta

operacional e tática. Brobst apresentou um modelo de cinco etapas, adequado à experiência de Coffee (2003) sobre como as organizações “crescem” em seu uso de dados (consulte Brobst et al.: 2005). Essas etapas (e as questões a que pretendem responder) são: relato (O que aconteceu?), análise (Por que aconteceu?), previsão (O que vai acontecer?), operacionalização (O que está acontecendo?) e *warehousing* ativo (O que eu quero que aconteça?). A última etapa, *warehousing* ativo, é onde podem ser obtidos os maiores benefícios. Muitas organizações estão aprimorando os *data warehouses* centralizados para que sirvam tanto a tomada de decisões operacionais quanto estratégicas.

Fontes: adaptado de P. Coffee, “‘Active’ Warehousing”, *eWEEK*, Vol. 20, No. 25, June 23, 2003, p. 36; and S. Brobst and C. Ballinger, *Active Data Warehousing*, Teradata Corp., [teradata.com/t/page/87127/October 2003](http://teradata.com/t/page/87127/October%202003) (acessado em abril de 2006).

investiu US\$ 600 milhões para usar dados e processos em tempo real. Esperava-se que o investimento cortasse 160 milhões de quilômetros nas entregas e economizasse 52 milhões de litros de combustível anualmente, através da gestão de suas tecnologias de fluxo de pacotes em tempo real (leia Malykhina: 2003). Na Tabela 2.4, mostramos uma comparação entre os ambientes de *data warehousing* tradicional e ativo.

Data warehousing em tempo real, *data warehousing* quase em tempo real, *warehousing* com latência zero e *data warehousing* ativo são nomes diferentes usados na prática para descrever o mesmo conceito. Gonzales (2005) apresentou definições diferentes para o ADW. Segundo o autor, ADW é apenas uma das opções que oferecem dados táticos e estratégicos misturados sob demanda. A arquitetura da construção de um ADW é muito parecida com a da fábrica de informações corporativa desenvolvida por Bill Inmon. A única diferença entre os dois métodos é a implemen-

Tabela 2.4 Comparação entre ambientes de data warehousing tradicional e ativo

Ambiente de data warehousing tradicional	Ambiente de data warehousing ativo
Apenas decisões estratégicas	Decisões estratégicas e táticas
Resultados às vezes difíceis de medir	Resultados medidos com operações
São aceitáveis atualizações de dados diária, semanal e mensal; resumos são normalmente adequados	São aceitáveis apenas dados abrangentes e detalhados, disponíveis em minutos
Simultaneidade moderada de usuários	Alto número de usuários (1.000 ou mais) que acessam e fazem consultas ao sistema simultaneamente
Relatórios altamente restritivos usados para confirmar ou verificar processos e padrões existentes; uso freqüente de tabelas de resumo ou data marts pré-desenvolvidos	Relatórios flexíveis ad hoc, assim como modelagem auxiliada por máquina (p. ex., data mining) para descobrir novas hipóteses e relações
Usuários avançados, trabalhadores da informação, usuários internos	Equipes operacionais, centrais de atendimento, usuários externos

Fontes: adaptado de P. Coffee, “‘Active’ Warehousing”, *eWEEK*, Vol. 20, No. 25, June 23, 2003, p. 36; and Teradata Corp., *Active Data Warehousing*, teradata.com/t/page/87127/index.html (acessado em abril de 2006).

tação dos dois *data stores* em um único ambiente. Entretanto, uma SOA baseada em XML e *Web services* oferece outra opção para misturar dados táticos e estratégicos sob demanda.

Uma questão crítica do *data warehousing* em tempo real é que nem todos os dados devem ser atualizados sempre. Isso causa problemas quando há geração de relatórios em tempo real, visto que os resultados de uma pessoa podem não coincidir com os de outra. Por exemplo, uma empresa que usa Business Objects Web Intelligence observou um problema importante com a inteligência em tempo real. Os relatórios em tempo real são completamente diversos quando produzidos em momentos diferentes (veja Peterson: 2003). Além do mais, pode não ser necessário atualizar constantemente certos dados, (por exemplo, conceitos de cursos de três ou mais anos atrás).

As exigências em tempo real mudam a forma como vemos o projeto de bancos de dados, *data warehouses*, OLAP e ferramentas de *data mining*, pois eles realmente são atualizados no mesmo momento em que há consultas ativas. Entretanto, é comprovado que agir desse modo tem seu valor significativo ao negócio, então é crucial que as organizações adotem esses métodos em seus processos. Nessas implementações, o planejamento cuidadoso é essencial.

Questões de revisão da Seção 2.7

1. O que é um RDW?
2. Liste os benefícios de um RDW.
3. Quais são as principais diferenças entre um *data warehouse* tradicional e um RDW?
4. Liste alguns dos elementos direcionadores do RDW.

2.8

QUESTÕES DE ADMINISTRAÇÃO E SEGURANÇA DE DATA WAREHOUSES

Os *data warehouses* oferecem uma vantagem competitiva distinta às empresas que os criam e utilizam eficazmente. Devido ao seu grande tamanho e à sua natureza intrínseca, o *data warehouse* exige um monitoramento particularmente forte a fim de manter um nível satisfatório de eficiência e produtividade. A administração e a gestão bem-sucedidas de um *warehouse* envolvem habilidades e proficiência que vão além do que se exige de um tradicional administrador de banco de dados (DBA). Um **administrador do data warehouse (DWA)** deve estar familiarizado com *hardware* e *software* de alta performance e com tecnologias de rede. Ele deve também ter uma sólida percepção dos negócios. Os *data warehouses* alimentam os DSS e os sistemas de BI que ajudam os gerentes a tomarem decisões; por isso, o DWA deve conhecer bem esses processos para projetar e manter a estrutura do sistema de forma adequada. É importante sobretudo que um DWA mantenha estáveis as exigências e recursos existentes do *warehouse*, ao mesmo tempo em que proporcione flexibilidade para rápidas melhorias. Por fim, um DWA deve ter habilidades excelentes de comunicação. Consulte Benander et al. (2000) para uma descrição das principais diferenças entre um DBA e um DWA.

A segurança e a privacidade das informações são uma preocupação principal e expressiva de um profissional do *data warehouse*. O governo norte-americano aprovou regulamentos (como as leis de privacidade e proteção Gramm Leach Bliley, ou o Health Insurance Portability and Accountability Act [HIPAA] de 1996) que instituíram exigências obrigatórias na gestão das informações do cliente. Assim, as empresas devem criar procedimentos de segurança que sejam eficientes e, ao mesmo tempo, flexíveis para estarem em conformidade com as diversas leis de privacidade. Segundo Elson e LeClerc (2005), um esquema eficiente de segurança em um *data warehouse* deve se focar em quatro áreas principais:

1. Estabelecimento de políticas e procedimentos corporativos e de segurança eficientes. Uma política de segurança eficiente deve começar no alto, com a gerência executiva, e ser comunicada a todas as pessoas da organização.
2. Implementação de procedimentos e técnicas lógicas de segurança para limitar o acesso. Isso inclui autenticação do usuário, controles de acesso e tecnologia de criptografia.

INSIGHTS DE TECNOLOGIA 2.8

Ambeo oferece solução comprovada de auditoria do acesso aos dados

Desde 1997, a Ambeo (ambeo.com; atual Embarcadero Technologies, Inc.) implementa tecnologias que oferecem gestão do desempenho, acompanhamento do uso dos dados, auditoria de privacidade dos dados e monitoramento para empresas Fortune 1000. Essas firmas possuem alguns dos maiores ambientes de banco de dados existentes. As soluções de auditoria do acesso aos dados da Ambeo desempenham um grande papel na infra-estrutura de segurança de informações de uma empresa.

A tecnologia da Ambeo é uma solução relativamente fácil que registra tudo o que acontece nos bancos de dados, com despesas gerais baixas ou nulas. Além disso, ela oferece auditoria do acesso aos dados que identifica exatamente quem olha os dados, quando olha, e o que está fazendo com os dados. Esse monitoramento em tempo real ajuda a identificar as falhas de segurança com rapidez e eficácia.

Fontes: adaptado de “Ambeo Delivers Proven Data Access Auditing Solution”, *Database Trends and Applications*, Vol. 19, No. 7, July 2005; and Ambeo, *Keeping Data Private (and Knowing It): Moving Beyond Conventional Safeguards to Ensure Data Privacy*, ambeo.com/why_ambeo_white_papers.html (acessado em abril de 2006).

3. Limitação do acesso físico ao ambiente do centro de dados.
4. Estabelecimento de um processo eficaz de revisão do controle interno, com ênfase em segurança e privacidade.

Leia nos *Insights* de tecnologia 2.8 uma descrição da importante ferramenta de *software* da Ambeo, que monitora a segurança e a privacidade dos *data warehouses*. Por fim, tenha em mente que acessar um *data warehouse* através de um dispositivo móvel sempre deve ser feito com cuidado. Neste caso, os dados devem apenas ter acesso somente leitura.

A curto prazo, os desenvolvimentos de *data warehousing* serão determinados por fatores perceptíveis (volumes de dados, maior tolerância a latência, a diversidade e a complexidade dos tipos de dados) e por fatores menos perceptíveis (requisitos do usuário final não-cumpridos, relativos a *dashboards*, *balanced scorecards*, gestão dos dados-mestre, qualidade das informações). Considerando-se esses motivadores, Agosta (2006) sugeriu que as tendências de *data warehousing* tomarão o rumo da simplicidade, do valor e do desempenho.

Questões de revisão da Seção 2.8

1. Quais passos uma organização pode seguir para garantir a segurança e a confidencialidade dos dados de clientes em seu *warehouse*?
2. Que habilidades um DWA deve ter? Por quê?

2.9

RECURSOS, LINKS E CONEXÃO COM A TERADATA UNIVERSITY NETWORK

O uso deste e da maioria dos outros capítulos do livro pode ser aprimorado pelas ferramentas descritas a seguir.

Recursos e links

Recomendamos os seguintes recursos e *links* para que você conheça outras leituras e explicações:

- The Data Warehousing Institute (tdwi.org)
- *DM Review* (dmreview.com)
- DSS Resources (dssresources.com)

Casos

Todos os principais fornecedores de sistemas de suporte ao gerenciamento (como MicroStrategy, Microsoft, Oracle, IBM, Hyperion, Cognos, Exsys, Fair Issac, SAP e Information Builders) têm histórias interessantes de sucesso com os clientes. Casos acadêmicos estão disponíveis em Harvard Business School Case Collection (harvardbusinessonline.hbsp.harvard.edu), Business Performance Improvement Resource (bpir.com), Idea Group Publishing (idea-group.com), Ivy League Publishing (ivylp.com), ICFAI Center for Management Research (icmr.icfai.org/casestudies/icmr_case_studies.htm), KnowledgeStorm (knowledgestorm.com) e em outros *sites*. Para recursos de caso adicionais consulte a Teradata University Network (teradatauniversitynetwork.com). Para casos de *data warehousing*, recomendamos especificamente os seguintes, provenientes da Teradata University Network (teradatauniversitynetwork.com): *Continental Airlines Flies High with Real-Time Business Intelligence*, *Data Warehouse Governance at Blue Cross and Blue Shield of North Carolina*, *3M Moves to a Customer Focus Using a Global Data Warehouse*, *Data Warehousing Supports Corporate Strategy at First American Corporation*, *Harrah's High Payoff from Customer Information e Whirlpool*. Recomendamos também a tarefa *Data Warehousing Failures*, que consiste em oito casos curtos de falhas de *data warehousing*.

Fornecedores, produtos e demos

Há uma lista abrangente em dmreview.com. Fornecedores são listados na Tabela 2.1. Veja também technologyevaluation.com.

Periódicos

Recomendamos os seguintes periódicos:

- *Baseline Magazine* (baselinemag.com)
- *Business Intelligence Journal* (tdwi.org)
- *CIO* (cio.com)
- *CIO Insight* (cioinsight.com)
- *Computer World* (computerworld.com)
- *Decision Support Systems* (elsevier.com)
- *DM Review* (dmreview.com)
- *eWEEK* (eweek.com)
- *InfoWeek* (infoworld.com)
- *InfoWorld* (infoworld.com)
- *InternetWeek* (internetweek.com)
- *Management Information Systems Quarterly (MIS Quarterly)* (misq.org)
- Revista online *Technology-Evaluation* (technologyevaluation.com)
- *Teradata Magazine* (teradata.com)

Referências adicionais

Para mais informações sobre *data warehousing*, consulte as fontes a seguir.

- Imhoff, C., N. Gallemmo, and J.G. Geiger. (2003). *Mastering Data Warehouse Design: Relational and Dimensional Techniques*. New York: Wiley.
- Marco, D., and M. Jennings. (2004). *Universal Meta Data Models*. New York: Wiley.
- Wang, J. (2005). *Encyclopedia of Data Warehousing and Mining*. Hershey, PA: Idea Group Publishing.

Para saber mais sobre bancos de dados, a estrutura sobre a qual são desenvolvidos os *data warehouses*, leia:

Watson, R.T. (2006). *Data Management*, 5th ed., New York:Wiley.

Conexão com a Teradata University Network (TUN)

A TUN (teradatauniversitynetwork.com) oferece uma riqueza de informações e casos sobre *data warehousing*. Um dos melhores é o caso da Continental Airlines, que você resolverá em um exercício mais adiante. Outros casos recomendados foram mencionados anteriormente neste capítulo. Na TUN, ao clicar na guia *Courses* e selecionar *Data Warehousing*, você verá *links* para muitos artigos, tarefas, capítulos de livros, *websites* de cursos, apresentações no PowerPoint, projetos, relatórios de pesquisa, planos de estudos e seminários na Web relacionados. Você encontrará também *links* para demonstrações de *softwares* de *data warehouses* ativos. Finalmente, você verá *links* para a Teradata (teradata.com), onde encontrará informações adicionais, incluindo excelentes histórias de sucesso de *data warehousing*, documentos informativos, cursos baseados na Web e a versão *online* da *Teradata Magazine*.

Destaques do capítulo

- Um *data warehouse* é um repositório de dados especialmente construído, no qual os dados são organizados para que os usuários finais facilmente os acessem com várias finalidades.
- Os *data marts* contêm dados sobre um tópico (por exemplo, *marketing*). Um *data mart* pode ser uma duplicação de um subconjunto de dados no *data warehouse*. Os *data marts* são uma solução menos cara, capaz de ser substituída por um *warehouse* ou de complementá-lo. Eles podem ser dependentes ou independentes de um *data warehouse*.
- Um *data store* operacional (ODS) é um tipo de banco de dados com arquivos de informações de clientes, usado freqüentemente como área de preparação para um *data warehouse*.
- A integração de dados compreende três grandes processos: acesso aos dados, federação de dados e captura de mudanças. Quando os três são implementados corretamente, os dados podem ser acessados e disponibilizados para uma gama de ferramentas de ETL e análise e de ambientes de *data warehousing*.
- As tecnologias de ETL extraem dados de muitas fontes, os limpam e carregam no *data warehouse*. ETL é um processo integral de qualquer projeto centrado em dados.
- O *data warehousing* ativo ou em tempo real complementa e amplia o *data warehousing*, entrando no âmbito da tomada de decisões operacionais e táticas ao carregar os dados em tempo real e oferecê-los aos usuários para uma tomada de decisões ativa.
- A segurança e a privacidade dos dados e informações é uma questão crítica para um profissional do *data warehouse*.

Termos-chave

- | | | |
|--|--|-------------------------|
| • administrador do <i>data warehouse</i> (DWA) | • <i>data warehouse</i> empresarial (EDW) | • metadados |
| • <i>data mart</i> | • <i>drill down</i> | • modelagem dimensional |
| • <i>data mart</i> dependente | • extração, transformação e carga (ETL) | • <i>oper mart</i> |
| • <i>data mart</i> independente | • grão | • tabela de dimensão |
| • <i>data store</i> operacional (ODS) | • integração de aplicações corporativas (EAI) | |
| • <i>data warehouse</i> | • integração de dados | |
| • <i>data warehouse</i> ativo (ADW) | • integração de informações corporativas (EII) | |
| • <i>data warehouse</i> em tempo real (RDW) | | |

QUESTÕES PARA DISCUSSÃO

1. Compare a integração de dados e a ETL. Como elas se relacionam?
2. O que é um data warehouse e quais são seus benefícios? Por que o acesso à Web é importante para um data warehouse?
3. Um data mart pode substituir ou complementar um data warehouse. Compare e discuta essas opções.
4. Discuta os principais motivadores e vantagens do data warehousing para os usuários finais.
5. Elabore uma lista com as diferenças e/ou semelhanças entre as funções de um administrador de um banco de dados e um administrador do data warehouse.
6. Descreva como a integração de dados pode levar a níveis mais altos de qualidade dos dados.
7. Compare as abordagens de Kimball e Inmon ao desenvolvimento de data warehouses. Identifique quando cada uma é mais eficiente.
8. Discuta as preocupações de segurança envolvidas na construção de um data warehouse.
9. Investigue a atual implementação do desenvolvimento de data warehouses através de terceirização no exterior. Escreva um relatório sobre o assunto. Em sala de aula, debata a questão em termos de custos, benefícios e fatores sociais.

Exercícios

Exercícios da Teradata University e outros exercícios interativos

1. Considere o caso que descreve o desenvolvimento e a aplicação de um data warehouse na Coca-Cola do Japão, disponível no website DSS Resources, dsresources.com/cases/coca-colajapan/index.html (há um resumo no Caso de aplicação 2.3). Leia o caso e responda às nove questões para aprofundar a análise e a discussão.
2. Leia o artigo de Ball (2006) e classifique os critérios (pensando em uma empresa real). Em um relatório, explique o quão importante é cada critério e por quê.
3. Explique quando se deve implementar uma arquitetura de duas ou três camadas ao considerar o desenvolvimento de um data warehouse.
4. Leia o caso completo da Continental Airlines em teradatastudentnetwork.com e responda às questões (o caso está resumido na vinheta de abertura).
5. Em teradatastudentnetwork.com, leia o caso “Harrah’s High Payoff from Customer Information” e responda às questões. Relacione os resultados da Harrah’s à forma como as companhias aéreas e outros cassinos usam os dados dos clientes.
6. Em teradatastudentnetwork.com, leia a tarefa “Data Warehousing Failures” e responda às questões. A tarefa descreve oito casos, então a turma pode ser dividida em oito grupos, cada um deles tratando de um caso. Leia também Ariyachandra e Watson (2006a) e, para cada caso, identifique como a falha ocorreu por não haver foco em um ou mais fatores de sucesso tomados como referência.
7. Em teradatastudentnetwork.com, leia e responda as questões da tarefa “AdVent Technology: Using the MicroStrategy Sales Analytic Model”. O software da MicroStrategy pode ser acessado pelo site da TUN. Recomendamos o uso da apresentação no PowerPoint de Barbara Wixom sobre o software da MicroStrategy (Demo Slides for MicroStrategy Tutorial Script), disponível no site da TUN.
8. Em teradatastudentnetwork.com, assista aos seminários na Web “Real-Time Data Warehousing: The Next Generation of Decision Support Data Management” e “Building the Real-Time Enterprise”. Leia o artigo “Teradata’s Real-Time Enterprise Reference Architecture: A Blueprint for the Future of IT”, disponível no mesmo website. Descreva como os conceitos e tecnologias de tempo real funcionam e de que maneira eles podem ser usados para ampliar as arquiteturas de BI e data warehousing existentes, a fim de apoiar a tomada de decisões do cotidiano. Escreva um relatório indicando como o data warehousing em tempo real especificamente fornece vantagem competitiva às organizações. Descreva detalhadamente as dificuldades da implementação e operação e como elas são trabalhadas na prática.
9. Em teradatastudentnetwork.com, assista aos seminários na Web “Data Integration Renaissance: New Drivers and Emerging Approaches”, “In Search of a Single Version of the Truth: Strategies for Consolidating Analytic Silos” e “Data Integration: Using ETL, EAI, and EII Tools to Create an Integrated Enterprise”. Leia também o relatório de pesquisa “Data integration”. Compare e diferencie as apresentações. Qual é a questão mais importante descrita nesses seminários? Qual é a melhor forma de lidar com as estratégias e os desafios de consolidar data marts e planilhas em uma arquitetura de data warehousing unificada? Realize uma pesquisa

na Web para identificar os últimos desenvolvimentos da área. Compare a apresentação ao material no texto e ao material novo que você encontrou.

10. Considere o futuro do *data warehousing*. Faça uma pesquisa na Internet sobre este tema. Leia também estes dois artigos: AGOSTA, L. "Data Warehousing in a Flat World: Trends for 2006". *DM Direct Newsletter*, 31 de março de 2006; GEIGER, J. G. "CIFE: Evolving with the Times". *DM Review*, novembro de 2005, pp. 38-41. Compare os artigos com suas descobertas.
11. Acesse teradatastudentnetwork.com. Identifique os mais recentes artigos, relatórios de pesquisa e casos sobre *data warehousing*. Descreva as últimas descobertas da área. Inclua em seu relatório como o *data warehousing* é usado em BI e DSS.

Tarefas em grupo e interpretação de papéis

1. Kathryn Avery trabalha há seis anos como DBA em uma rede varejista (Big Chain) conhecida em todo o país. Recentemente, foi solicitado a ela que liderasse o desenvolvimento do primeiro *data warehouse* da Big Chain. O projeto era patrocinado pela alta gerência e pelo CIO. A justificativa para o desenvolvimento do *warehouse* é aprimorar os sistemas de relatórios, principalmente em vendas e *marketing*, e, a longo prazo, melhorar o CRM da Big Chain. Kathryn foi a uma conferência do Data Warehousing Institute e está fazendo algumas leituras, mas ainda está incerta quanto as metodologias de desenvolvimento do sistema. Ela sabe que há dois grupos de arquitetura – EDW (Inmon) e *data marts* (Kimball) – em iguais condições de fornecer serviços robustos.

Inicialmente, ela acreditava que as duas metodologias eram extremamente desiguais; porém, após examiná-las mais atentamente, já não tem tanta certeza. Kathryn tem algumas questões que gostaria de ter respondidas:

- a. Quais são as diferenças reais entre as metodologias?
- b. Que fatores contam na hora de selecionar uma determinada metodologia?
- c. Quais devem ser seus próximos passos ao pensar sobre uma metodologia?

Ajude Kathryn a responder a essas perguntas. (Este exercício foi adaptado de K. Duncan, L. Reeves, and J. Griffin, "BI Experts' Perspective", *Business Intelligence Journal*, Vol. 8, No. 4, Fall 2003, pp. 14–19.)

2. Jeet Kumar é o administrador do *data warehousing* de um grande banco regional. Há cinco anos, ele foi selecionado para implementar um *data warehouse* em suporte à estratégia de negócios de CRM do banco. Usando o *warehouse*, o banco teve sucesso

ao integrar as informações dos clientes, compreender sua lucratividade, atrair e manter clientes e melhorar suas relações com eles.

Com o passar dos anos, o *data warehouse* do banco se aproximou das operações em tempo real, pois passou a atualizar com mais frequência os dados no *warehouse*. Agora, o banco quer implementar aplicativos de auto-atendimento e de central de atendimento aos clientes, o que exige dados ainda mais atualizados do que os agora disponíveis no *warehouse*.

Jeet quer um auxílio para considerar as possibilidades de obtenção de dados mais recentes. Uma alternativa é comprometer-se totalmente com a implementação do *data warehousing* em tempo real. Seu fornecedor de ETL está preparado para ajudá-lo nesta mudança. Mesmo assim, Jeet se informou sobre as tecnologias de EAI e EII e pensa em como elas podem se encaixar em seus planos.

Suas dúvidas são principalmente as seguintes:

- a. O que exatamente são as tecnologias de EAI e EII?
- b. Como a EAI e EII se relacionam a ETL?
- c. Como a EAI e EII se relacionam ao *data warehousing* em tempo real?
- d. EAI e EII são necessários, complementares ou alternativas ao *data warehousing* em tempo real?

Ajude Jeet a responder a essas perguntas. (Este exercício foi adaptado de S. Brobst, E. Levy, and C. Muzilla, "Enterprise Application Integration and Enterprise Information Integration", *Business Intelligence Journal*, Vol. 10, No. 2, Spring 2005, pp. 27–32.)

3. Entreviste administradores de sua instituição de ensino ou executivos na empresa em que você trabalha para determinar como o *data warehousing* poderia auxiliá-los em seu trabalho. Redija uma proposta descrevendo suas descobertas. Inclua as estimativas de custo e os benefícios.
4. Percorra a lista dos riscos de *data warehousing* descritos neste capítulo e encontre dois exemplos práticos de cada um.
5. Acesse teradata.com e leia os documentos informativos "Measuring Data Warehouse ROI" e "Realizing ROI: Projecting and Harvesting the Business Value of an Enterprise Data Warehouse". Assista também ao curso baseado na web "The ROI Factor: How Leading Practitioners Deal with the Tough Issue of Measuring DW ROI". Descreva as questões mais importantes contidas nos materiais. Compare-as aos fatores de sucesso descritos por Ariyachandra e Watson (2006a).
6. Leia o artigo: K. Liddell Avery and Hugh J. Watson, "Training Data Warehouse End-users", *Business*

Intelligence Journal, Vol. 9, No. 4, Fall 2004, pp. 40–51 (disponível em **teradatastudent.network.com**). Considere as diferentes classes de usuários finais, descreva suas dificuldades e apresente os benefícios do treinamento adequado para cada grupo. Combine que cada membro do grupo assuma um dos papéis, e que o grupo discuta sobre como determinado tipo de treinamento em *data warehousing* seria bom para cada um.

Exercícios na Internet

1. Faça uma busca na Internet por informações sobre *data warehousing*. Identifique *newsgroups* que tenham interesse neste conceito. Explore os recursos da ABI/INFORM em sua biblioteca física ou virtual e use o Google para pesquisar artigos recentes sobre o tema. Comece com **tdwi.org**, **technologyevaluation.com**, e os principais fornecedores: **teradata.com**, **sas.com**, **oracle.com** e **ncr.com**. Confira também **cio.com**, **dmreview.com**, **dssresources.com** e **db2mag.com**.
2. Examine algumas ferramentas e fornecedores de ETL. Comece com **fairisaac.com** e **egain.com**. Consulte ainda **dmreview.com**.
3. Entre em contato com alguns fornecedores de *data warehouse* e consiga informações sobre seus produtos. Dê atenção especial aos que oferecem ferramentas para vários propósitos, como Cognos, Software A&G, SAS Institute e Oracle. Versões *demo online* gratuitas estão disponíveis em alguns desses fornecedores. Faça o *download* de um *demo* ou dois e experimente. Escreva um relatório sobre sua experiência.
4. Explore **teradata.com** e busque desenvolvimentos e histórias de sucesso de *data warehousing*. Elabore um relatório sobre o que você descobriu.
5. Procure documentos informativos e cursos baseados na Web sobre *data warehousing* no site **teradata.com**. Leia os documentos e assista aos cursos (divida a turma para cobrir todas as fontes). Escreva em um relatório o que você encontrou.
6. Encontre casos recentes de aplicações bem-sucedidas de *data warehousing*. Vá em *websites* de fornecedores de *data warehouses* e procure casos ou histórias de sucesso. Selecione um deles e redija um breve resumo para apresentar à turma.

CASO DE APLICAÇÃO DE FINAL DO CAPÍTULO

Data warehousing em tempo real na Overstock.com

Na primavera de 1999, o Dr. Patrick M. Byrne reconheceu o potencial de liquidar estoques excedentes na Internet. Seis meses depois e sem financiamento externo, ele lançou o *website* Overstock.com. A Overstock.com, Inc., é uma varejista de ponta de estoque *online* que vende mercadorias de marca com desconto na Internet. A empresa oferece a seus clientes uma oportunidade conveniente de comprar saldos e a seus fornecedores um canal alternativo de distribuição para liquidar o estoque. Criada sobre os princípios de investimento de valor e negociação justa, a Overstock.com tornou-se rapidamente a líder *online* de um mercado avaliado em US\$ 60 bilhões nos Estados Unidos. O número de produtos que a loja oferece cresceu de menos de 100, em 1999, a mais de 14.000 itens – além dos aproximadamente 650.000 produtos de livros, música e vídeo – em junho de 2005. A Overstock.com oferece os melhores valores em produtos de marca para os consumidores.

Em abril de 2005, a Overstock.com selecionou a Teradata como provedora de EDW com aplicações

analíticas, incluindo CRM, *data mining*, BA, a suíte de *e-business* da Teradata, modelos de dados lógicos e consultoria. Este EDW integrou os dados da empresa em um repositório único para proporcionar uma visão holística dos negócios. O EDW da Teradata oferece uma tecnologia de banco de dados paralelo, uma suíte de utilitários de acesso e gestão de dados e um portfólio de ferramentas de análise que apresentam insights detalhados de negócios, além de promoverem interações com um alto grau de relevância para as necessidades dos clientes.

Em agosto de 2005, a loja, sediada em Salt Lake City, começou a conectar os usuários a um *data warehouse* em tempo real. Ela usou ferramentas de gestão de dados transacionais da GoldenGate Software, Inc. para extrair as informações diretamente dos seus sistemas comerciais para o *warehouse*. A Overstock.com adotou também a estrutura de serviços baseada na Web da Teradata para a integração de aplicações, visando substituir um processo que usava ferramentas de ETL tradicionais para elaborar relatórios diretamente

a partir dos seus sistemas de retaguarda. Agora, o *data warehouse* recebe os dados de acesso ao *website* em tempo real, dados financeiros e de vendas de produtos a cada 15 minutos e outras informações de hora em hora. “Quando lançamos alguma campanha, podemos saber dentro de 15 minutos se elas estão produzindo uma alta na receita que normalmente não aconteceria”, diz Jack Garcella, vice-presidente de análise e relatórios de *data warehousing*. Com alguns terabytes de dados armazenados no EDW e milhões de transações efetivadas todos os dias, a Overstock.com precisava garantir que o *warehouse* fosse atualizado constantemente. Ela selecionou o Sunopsis Data Conductor, um forte produto de ETL, para permitir o trabalho automático e eficiente com grandes volumes de dados provenientes dos sistemas de origem.

Fontes: adaptado de K. Ferrell, “Mission Possible: An Ambitious Data Warehouse Implementation—and Great Results—Are All in a Day’s Work for Overstock.com”, *Teradata Magazine*, Vol. 6, No. 1, March 2006; J. Garzella, “Overstock.com Meets High-Volume, High-Performance Real-Time Data Transformation Needs with Sunopsis”, *DM Review*, Ja-

nuary 2006; C. Babcock, “Overstock.com Moves Reports to Data Warehouse”, *InformationWeek*, August 3, 2005; *History, Overstock.com Investor Relations, shareholder.com* (acessado em abril de 2006); **Shareholder.com** Investor Relations, *Overstock.com, Inc. Selects Teradata for Enterprise Data Warehouse, Customer Analytics and Single View of the Business, shareholder.com* (acessado em abril de 2006); and Teradata.com, *Overstock.com, Inc. Selects Teradata for Enterprise Data Warehouse, Customer Analytics and Single View of the Business*, April 20, 2005, teradata.com/t/go.aspx/index.html?id=133353 (acessado em abril de 2006).

QUESTÕES SOBRE O CASO

1. Descreva como a Overstock.com executava suas campanhas antes do sistema da Teradata ser desenvolvido.
2. Como o *data warehouse* em tempo real ajudou a Overstock.com a melhorar o desempenho dos negócios?
3. Visite os *websites* dos fornecedores citados no caso e examine os recursos e capacidades de *data warehousing* de cada um. Descreva detalhadamente como a Overstock.com poderia usar cada um deles.

Referências

- Adelman, S., and L. Moss. (2001, Winter). “Data Warehouse Risks.” *Journal of Data Warehousing*, Vol. 6, No. 1.
- Agosta, L. (2006, January). “The Data Strategy Adviser: The Year Ahead—Data Warehousing Trends 2006.” *DM Review*, Vol. 16, No. 1.
- Akbay, S. (2006, Quarter 1). “Data Warehousing in Real Time.” *Business Intelligence Journal*, Vol. 11, No. 1.
- Anthes, G.H. (2003, June 30). “Hilton Checks into New Suite.” *Computerworld*, Vol. 37, No. 26.
- Ariyachandra, T., and H.Watson. (2005). “Key Factors in Selecting a Data Warehouse Architecture.” *Business Intelligence Journal*, Vol. 10, No. 2.
- Ariyachandra, T., and H.Watson. (2006a, January). “Benchmarks for BI and Data Warehousing Success.” *DM Review*, Vol. 16, No. 1.
- Ariyachandra, T., and H.Watson. (2006b, Quarter 1). “Which Data Warehouse Architecture Is Most Successful?” *Business Intelligence Journal*, Vol. 11, No. 1.
- Armstrong, R. (2000, Quarter 3). “E-nalysis for the E-business”, *Teradata Magazine Online*, teradata.com.
- Ball, S.K. (2005, November 14). *Do You Need a Data Warehouse Layer in Your Business Intelligence Architecture?* datawarehouse.ittoolbox.com/documents/industry-articles/do-you-need-a-datawarehouse-layer-in-your-business-intelligencearchitecture-2729 (acessado em abril 2006).
- Barquin, R., A. Paller, and H. Edelstein. (1997). “Ten Mistakes to Avoid for Data Warehousing Managers.” In R. Barquin and H. Edelstein (eds.). *Building, Using, and Managing the Data Warehouse*. Upper Saddle River, NJ: Prentice Hall.
- Basu, R. (2003, November). “Challenges of Real-Time Data Warehousing.” *DM Review*.
- Bell, L.D. (2001, Spring). “MetaBusiness Meta Data for the Masses: Administering Knowledge Sharing for Your Data Warehouse.” *Journal of Data Warehousing*, Vol. 6, No. 2.
- Benander, A., B. Benander, A. Fadlalla, and G. James. (2000, Winter). “Data Warehouse Administration and Management”, *Information Systems Management*, Vol. 17, No. 1.
- Bonde, A., and M. Kuckuk. (2004, April). “Real World Business Intelligence: The Implementation Perspective.” *DM Review*, Vol. 14, No. 4.
- Breslin, M. (2004, Winter). “Data Warehousing Battle of the Giants: Comparing the Basics of Kimball and Inmon Models.” *Business Intelligence Journal*, Vol. 9, No. 1.
- Brobst, S., E. Levy, and C. Muzilla. (2005, Spring). “Enterprise Application Integration and Enterprise Information Integration.” *Business Intelligence Journal*, Vol. 10, No. 2.
- Brody, R. (2003, Summer). “Information Ethics in the Design and Use of Metadata.” *IEEE Technology and Society Magazine*, Vol. 22, No. 2.

- Brown, M. (2004, May 9–12). “8 Characteristics of a Successful Data Warehouse.” *Proceedings of the Twenty-Ninth Annual SAS Users Group International Conference (SUGI 29)*. Montreal.
- Burdett, J., and S. Singh. (2004). “Challenges and Lessons Learned from Real-Time Data Warehousing.” *Business Intelligence Journal*, Vol. 9, No. 4.
- Coffee, P. (2003, June 23). “‘Active’ Warehousing.” *eWE-EK*, Vol. 20, No. 25.
- Dasu, T., and T. Johnson. (2003). *Exploratory Data Mining and Data Cleaning*. New York: Wiley.
- Davison, D. (2003, November 14). “Top 10 Risks of Offshore Outsourcing.” META Group Research Report, now Gartner, Inc., Stamford, CT.
- Devlin, B. (2003, Quarter 2). “Solving the Data Warehouse Puzzle.” *DB2 Magazine*.
- Dragoon, A. (2003, July 1). “All for One View.” *CIO*.
- Eckerson, W. (2003, Fall). “The Evolution of ETL.” *Business Intelligence Journal*, Vol. 8, No. 4.
- Eckerson, W. (2004, November) “Gauge Your Warehouse Maturity.” *DM Review*, Vol. 14, No. 11.
- Eckerson, W. (2005, April 1). “Data Warehouse Builders Advocate for Different Architectures.” *Application Development Trends*.
- Edwards, M. (2003, Fall). “2003 Best Practices Awards Winners: Innovators in Business Intelligence and Data Warehousing.” *Business Intelligence Journal*, Vol. 8, No. 4.
- Elson, R., and R. LeClerc. (2005). “Security and Privacy Concerns in the Data Warehouse Environment.” *Business Intelligence Journal*, Vol. 10, No. 3.
- Ericson, J. (2006, March). “Real-Time Realities.” *BI Review*.
- Gonzales, M. (2005, Quarter 1). “Active Data Warehouses Are Just One Approach for Combining Strategic and Technical Data.” *DB2 Magazine*.
- Hall, M. (2002, April 15). “Seeding for Data Growth.” *Computerworld*, Vol. 36, No. 16.
- Hicks, M. (2001, November 26). “Getting Pricing Just Right.” *eWEEK*, Vol. 18, No. 46.
- Hoffer, J.A., M.B. Prescott, and F.R. McFadden. (2007). *Modern Database Management*, 8th ed. Upper Saddle River, NJ: Prentice Hall.
- Hwang, M., and H. Xu. (2005, Fall). “A Survey of Data Warehousing Success Issues.” *Business Intelligence Journal*, Vol. 10, No. 4.
- Imhoff, C. (2001, May). “Power Up Your Enterprise Portal.” *E-Business Advisor*.
- Inmon, W.H. (2005). *Building the Data Warehouse*, 4th ed. New York: Wiley.
- Inmon, W.H. (2006, January). “Information Management: How Do You Tune a Data Warehouse?” *DM Review*, Vol. 16, No. 1.
- Jukic, N., and C. Lang. (2004, Summer). “Using Offshore Resources to Develop and Support Data Warehousing Applications.” *Business Intelligence Journal*, Vol. 9, No. 3.
- Karacsony, K. (2006, January). “ETL Is a Symptom of the Problem, not the Solution.” *DM Review*, Vol. 16, No. 1.
- Kassam, S. (2002, April 16). “Freedom of Information.” *Intelligent Enterprise*, Vol. 5, No. 7.
- Kay, R. (2005, September 19). “EII.” *Computerworld*. Vol. 39, No. 38.
- Kelly, C. (2001, June 14). “Calculating Data Warehousing ROI.” *SearchSQLServer.com Tips*.
- Malykhina, E. (2003, January 3). “The Real-Time Imperative.” *InformationWeek*, Issue 1020.
- Manglik, A., and V. Mehra. (2005, Winter). “Extending Enterprise BI Capabilities: New Patterns for Data Integration.” *Business Intelligence Journal*, Vol. 10, No. 1.
- Matney, D. (2003, Spring). “End-User Support Strategy.” *Business Intelligence Journal*, Vol. 8, No. 2.
- McCloskey, D.W. (2002). *Choosing Vendors and Products to Maximize Data Warehousing Success*. Auerbach Publications, auerbach-publications.com.
- Mehra, V. (2005, Summer). “Building a Metadata-Driven Enterprise: A Holistic Approach.” *Business Intelligence Journal*, Vol. 10, No. 3.
- Murtaza, A. (1998, Fall). “A Framework for Developing Enterprise Data Warehouses.” *Information Systems Management*, Vol. 15, No. 4.
- Nash, K.S. (2002, July). “Chemical Reaction.” *Baseline*. Orovic, V. (2003, June). “To Do & Not to Do.” *eAI Journal*.
- Parzinger, M.J., and M.N. Frolick. (2001, July). “Creating Competitive Advantage Through Data Warehousing.” *Information Strategy*, Vol. 17, No. 4.
- Peterson, T. (2003, April 21). “Getting Real About Real Time.” *ComputerWorld*, Vol. 37, No. 16.
- Raden, N. (2003, June 30). “Real Time: Get Real, Part II.” *Intelligent Enterprise*.
- Rosenberg, A. (2006, Quarter 1). “Improving Query Performance in Data Warehouses.” *Business Intelligence Journal*, Vol. 11, No. 1.
- Sammon, D., and P. Finnegan. (2000, Fall). “The Ten Commandments of Data Warehousing.” *Database for Advances in Information Systems*, Vol. 31, No. 4.
- Sapir, D. (2005, May). “Data Integration: A Tutorial.” *DM Review*, Vol. 15, No. 5.
- Sen, A. (2004, April). “Metadata Management: Past, Present and Future.” *Decision Support Systems*, Vol. 37, No. 1.
- Sen, A., and P. Sinha (2005). “A Comparison of Data Warehousing Methodologies.” *Communications of the ACM*, Vol. 48, No. 3.
- Solomon, M. (2005, Winter). “Ensuring a Successful Data Warehouse Initiative.” *Information Systems Management Journal*.
- Songini, M.L. (2004, February 2). “ETL Quickstudy.” *Computerworld*, Vol. 38, No. 5.
- Tannenbaum, A. (2002, Spring). “Identifying Meta Data Requirements.” *Journal of Data Warehousing*, Vol. 7, No. 2.
- Tennant, R. (2002, May 15). “The Importance of Being Granular.” *Library Journal*, Vol. 127, No. 9.
- Terr, S. (2004, February). “Real-Time Data Warehousing: Hardware and Software.” *DM Review*, Vol. 14, No. 2.

- Turban, E., D. Leidner, E. McLean, and J. Wetherbe. (2006). *Information Technology for Management*, 5th ed. New York: Wiley.
- Vaduva, A., and T. Vetterli. (2001, September). "Metadata Management for Data Warehousing: An Overview." *International Journal of Cooperative Information Systems*, Vol. 10, No. 3.
- Van den Hoven, J. (1998). "Data Marts: Plan Big, Build Small." *Information Systems Management*, Vol. 15, No. 1.
- Watson, H.J. (2002). "Recent Developments in Data Warehousing." *Communications of the ACM*, Vol. 8, No. 1.
- Watson, H., J. Gerard, L. Gonzalez, M. Haywood, and D. Fenton. (1999), "Data Warehouse Failures: Case Studies and Findings." *Journal of Data Warehousing*, Vol. 4, No. 1.
- Watson, H.J., D.L. Goodhue, and B.H. Wixom. (2002). "The Benefits of Data Warehousing: Why Some Organizations Realize Exceptional Payoffs." *Information & Management*, Vol. 39.
- Weir, R. (2002, Winter). "Best Practices for Implementing a Data Warehouse." *Journal of Data Warehousing*, Vol. 7, No. 1.
- Wilk, L. (2003, Spring). "Data Warehousing and Real-Time Computing." *Business Intelligence Journal*, Vol. 8, No. 2.
- Wixom, B., and H. Watson. (2001, March). "An Empirical Investigation of the Factors Affecting Data Warehousing Success." *MIS Quarterly*, Vol. 25, No. 1.
- Zhao, X. (2005, October 7). "Meta Data Management Maturity Model", *DM Direct Newsletter*.



Análise de Negócios e Visualização de Dados

Objetivos de aprendizado

- ◆ Descrever a análise de negócios (BA) e sua importância para as organizações.
- ◆ Listar e descrever brevemente os principais métodos e ferramentas de BA.
- ◆ Descrever como o processamento analítico *online* (OLAP), a visualização de dados e a multidimensionalidade podem melhorar a tomada de decisões.
- ◆ Descrever métodos de análise avançada.
- ◆ Descrever os sistemas de informações geográficas (SIG) e seu apoio à tomada de decisões.
- ◆ Descrever a BA em tempo real.
- ◆ Descrever como a *business intelligence* (BI) dá suporte à inteligência competitiva.
- ◆ Descrever os sistemas de suporte automatizado à decisão (ADS) e seus benefícios.
- ◆ Explicar como a Web se relaciona à BA.
- ◆ Descrever a inteligência da Web e a análise da Web e sua importância para as organizações.
- ◆ Descrever questões de implementação relacionadas à BA e os fatores de sucesso em BA.

Muitas organizações acumularam enormes quantidades de dados que podem ser usados pelos funcionários para desvendar relações preciosas para permitir que a organização tenha sucesso ao competir com outras. Algumas empresas fazem isso muito bem, usando ferramentas de análise; outras já não têm tanta eficiência. O ato de permitir a análise da decisão por meio de acesso a todos os dados e informações relevantes é conhecido como *análise de negócios (BA)*. A BA inclui OLAP, multidimensionalidade, visualização de dados, SIG, *data mining* e técnicas de análise avançada. Os dados de entrada analisados com o uso de BA podem ser encontrados em um *data warehouse*, ser dados operacionais, ou ainda estar em arquivos da Internet. Realizar esta análise não é tarefa simples e pode não dar certo se não for feito corretamente. Este capítulo descreve essas questões nas seguintes seções:

- 3.1 Vinheta de abertura: a Lexmark International aperfeiçoa as operações com o *business intelligence*
- 3.2 O campo da análise de negócios (BA): uma visão geral
- 3.3 Processamento analítico *online* (OLAP)
- 3.4 Relatórios e consultas
- 3.5 Multidimensionalidade
- 3.6 Análise de negócios avançada

- 3.7 Visualização de dados
- 3.8 Sistemas de informações geográficas (SIG)
- 3.9 *Business intelligence* em tempo real, suporte automatizado à decisão (ADS) e inteligência competitiva
- 3.10 Análise de negócios e a Web: inteligência da Web e análise da Web
- 3.11 Uso, benefícios e sucesso da análise de negócios

3.1

VINHETA DE ABERTURA: A LEXMARK INTERNATIONAL APERFEIÇOAS AS OPERAÇÕES COM O *BUSINESS INTELLIGENCE*

A Lexmark International (**lexmark.com**) é um fabricante global de produtos e soluções de impressão. Ela conta com cerca de 12 mil funcionários e mais de 50 escritórios de vendas em todo o mundo. Milhares de parceiros comerciais vendem os produtos da Lexmark em mais de 160 países.

Problema

Por estar em um setor extremamente competitivo, a Lexmark necessita de informações detalhadas, precisas e oportunas para o suporte à decisão e a implementação de estratégias. Isso ganha uma importância particular quando se trata do fluxo de dados entre a Lexmark e seus parceiros no varejo. As informações mais expressivas são sobre volumes de vendas em todos os lugares e níveis de estoque. O sistema antigo era lento, ineficiente e cheio de erros. Os problemas ocorriam tanto com fluxos provenientes dos parceiros quanto com o fornecimento de dados. Após o fornecimento dos dados, os resultados com frequência eram copiados de planilhas e colados em relatórios; o processo normalmente levava quatro dias ou mais para produzir respostas a perguntas comuns de negócios. Os representantes de vendas que trabalhavam externamente tinham de conectar-se à *intranet*. Uma vez estabelecida a conexão, os analistas e os representantes de vendas tinham de escrever consultas em *Structured Query Language* (SQL) e navegar pelo *mainframe* para gerar relatórios à gerência, alguns dos quais eram baseados em dados imprecisos, de uma semana atrás.

Solução

A Lexmark implementou uma solução de BI da MicroStrategy. A aplicação é uma adaptação de BI para o varejo, conhecida hoje como Retail BI System. Este sistema permite que compradores, analistas financeiros, analistas de *marketing*, gerentes regionais, comerciantes e representantes de vendas externos analisem os dados de vendas e estoques a partir de seus computadores ou dispositivos móveis. O sistema, alimentado por *data warehouse* da IBM, oferece aos usuários a capacidade de acompanhar o desempenho de vendas e os níveis de estoque de cada produto Lexmark, em milhares de lojas espalhadas pelo mundo. O *software* contém um grande número de ferramentas de relatório e análise, incluindo relatórios com mais de 50 modelos estatísticos e técnicas de visualização (descritas posteriormente neste capítulo). Ao utilizar o sistema, a comunidade de usuários da Lexmark pode responder instantânea e facilmente a consultas como estas:

- Quais são os níveis semanais de vendas e estoque em cada uma das lojas de um cliente específico pelo país?
- Quem foram os principais vendedores de determinado produto no varejo na semana passada, no mês passado ou neste final de semana comparado ao final de semana passado?
- Considerando uma dada loja que repassa à Lexmark dados de vendas e estoque por intercâmbio eletrônico de dados (EDI), quais são os níveis de estoque de um determinado produto campeão de vendas?

Resultados

A Lexmark informou que os tomadores de decisão agora têm informações oportunas, precisas e detalhadas. O novo sistema ajudou a identificar oportunidades de vendas, aumentou a fidelidade dos parceiros, eliminou problemas de estoque e fez a lucratividade crescer. Por exemplo, quando a empresa identifica que um determinado ponto de varejo está prestes a vender suas últimas impressoras de um modelo, um *alerta* automático é enviado ao gerente da loja e, dentro de algumas horas, é feito um pedido de reposição, o que evita a falta de estoque. No geral, foram recuperadas 100 mil vendas potencialmente perdidas. As lojas gostam da Lexmark devido a este serviço, o que a torna um fornecedor preferido.

Quase todos os funcionários da Lexmark usam este serviço. Os trabalhadores novos usam as informações para melhorar a forma como trabalham. Os gerentes podem entender melhor as tendências de negócios e tomar decisões estratégicas apropriadas. Eles têm um entendimento mais completo das demandas do consumidor por país ou loja, e, assim, podem decidir melhor sobre aspectos como precificação e promoções. Além disso, os serviços aos clientes e aos parceiros tiveram uma grande melhoria.

Fontes: compilado de MicroStrategy, *Success Story: Lexmark*, 2006, microstrategy.com/Customers/Successes/lexmark.asp (acessado em fevereiro de 2006); and L.Valentine, “Lexmark CIO Crowell P. Chambers: Supporting a Changing Business Environment”, *CIO Today*, July 21, 2004, ciotoday.com/story.xhtml?story_id=25966 (acessado em março de 2006).

Questões sobre a vinheta de abertura

1. Identifique os desafios de fluxo de informações que a Lexmark enfrentou.
2. Como os fluxos de informação eram fornecidos antes e como são fornecidos depois da implementação do sistema?
3. Identifique as decisões que recebem suporte do novo sistema.
4. Como o novo sistema pode aperfeiçoar o serviço ao cliente?
5. Visite microstrategy.com. Examine os recursos do Retail Business Intelligence System. Prepare uma lista dos recursos.
6. Entre no *site* sas.com, encontre o produto Retail Intelligence e faça o *tour* interativo. Compare-o ao produto da MicroStrategy. Compare-o também ao Oracle Retail (veja em oracle.com/applications/retail.html).

O que podemos aprender com esta vinheta?

O caso da Lexmark indica a necessidade de realizar coleta e análise das informações, que podem estar em centenas de lugares, e de fazê-lo de forma pontual. As informações também precisam ser distribuídas rapidamente. Isso pode ser conseguido das seguintes maneiras:

- Criação de um *data warehouse* que integra e padroniza dados provenientes de muitas fontes e locais
- Uso de *software* que permite fácil consulta dos dados de qualquer local, a qualquer momento.
- Uso de *software* para preparar e distribuir relatórios
- Uso de *software* de BI para realizar várias análises nos dados, visando descobrir oportunidades ou problemas e encontrar soluções e tendências
- Uso de *software* de BI para alertar gerentes e outros
- Descoberta do status de vendas e estoques de forma rápida e precisa (ou seja, usando **relatórios de status**)

Esses recursos são comuns nas atividades realizadas com *softwares* de análise de BI e são o assunto deste capítulo. São chamados de *análise de negócios*.

O CAMPO DA ANÁLISE DE NEGÓCIOS (BA): VISÃO GERAL

Conforme descrito nos Capítulos 1 e 5, a **business intelligence (BI)** implica em adquirir dados e informações (e, talvez, conhecimento) de uma grande variedade de fontes, organizá-los em um *data warehouse* e usá-los na tomada de decisões. A BA oferece os modelos e procedimentos de análise para a BI. O processo também envolve acompanhar os dados e analisá-los para obter vantagem competitiva. Para maiores detalhes, confira Eckerson (2003). Vejamos agora como a BA funciona.

Os fundamentos da BA

Análise é o método usado pela ciência analítica. Normalmente, ela se refere à análise de dados. Há muitas formas de realizar uma análise. De fato, existem muitos métodos e centenas de ferramentas de *software* para conduzir análises.

A **análise de negócios (BA)** é uma ampla categoria de aplicações e técnicas para reunir, armazenar, analisar e fornecer acesso aos dados, com o objetivo de ajudar os usuários da empresa a tomarem melhores decisões comerciais e estratégicas. A BA é conhecida também como *processamento analítico*, *ferramentas de BI*, *aplicações de BI* e simplesmente *BI*. (Nota: consulte DEVLIN: 2006 para um glossário com esses e outros termos relacionados.) A BI está se tornando uma ferramenta essencial para a maioria das empresas médias e grandes. A Pizza Hut, por exemplo, impulsionou suas vendas de forma significativa usando ferramentas de BI (veja LANGNAU: 2003). Baseada no valor de 20 anos de dados sobre consumidores, a Pizza Hut sabe que tipo de pizzas os clientes pedem, que tipos de cupons normalmente usam e quanto gastam em um dado período. Os gerentes de *marketing* podem operar essas informações através de uma análise de BI que prevê, por exemplo, a probabilidade do próximo pedido de um cliente. A empresa então usa essa informação para determinar estratégias de *marketing* para influenciar o cliente a comprar mais pizzas, sem gastar mais que o necessário na estratégia. O processo de BI geralmente, mas não necessariamente, envolve o uso de um *data warehouse*, ou até mesmo a construção de um.

Uma aplicação analítica é um passo avante rumo à sofisticação, em comparação ao simples oferecimento de técnicas ou ferramentas de análise. Ela permite atividades como:

- Automatização do pensamento e, na maioria dos casos, de uma parte da tomada de decisões de um ser humano.
- Uso comum de técnicas quantitativas complexas, como análise de regressão multivariada, *data mining*, inteligência artificial ou programação não-linear.

Exemplo

Uma aplicação analítica usada na pontuação de crédito de um candidato a empréstimo pode:

- Calcular a pontuação da capacidade de crédito.
 - Aceitar ou rejeitar automaticamente a solicitação de empréstimo.
 - Selecionar o limite de empréstimo.
 - Selecionar que produto do cartão de crédito (taxa de juros, termos de pagamento etc.) emitir a este candidato, ou qual outro tipo de empréstimo aprovar.
-

Ao usar *software* para BA, o usuário pode fazer consultas, requisitar relatórios *ad hoc* ou realizar análises. Por exemplo, é possível fazer análises executando consultas em várias camadas. Todos os bancos de dados estão vinculados, e por isso se consegue pesquisar quais produtos estão sobrando no estoque de uma determinada loja. Você pode determinar quais desses produtos normalmente são vendidos com itens populares, baseado em vendas anteriores. Após planejar uma promoção para unir o excesso de estoque com os produtos populares (como em um pacote de produtos), é possível explorar os dados mais a fundo para ver onde esta promoção seria mais popular

(e mais rentável). Os resultados de sua solicitação podem vir na forma de relatórios, previsões, alertas e/ou apresentações gráficas, os quais são difundidos entre os tomadores de decisão. Para ver um exemplo de aplicação na Ben & Jerry's, leia o Caso de aplicação 3.1.

CASO DE APLICAÇÃO 3.1

Ben & Jerry's se supera com a BA

Na fábrica da Ben & Jerry's (benjerry.com) em Waterbury, Vermont, tubos enormes bombeiam mais de 94.000 litros de sorvete todos os dias. Durante o dia, caminhões refrigerados fazem fila, recolhem o sorvete e o distribuem aos depósitos. Dali, o sorvete é enviado para mais de 60.000 mercados nos Estados Unidos e em outros 14 países. Nos mercados, o sorvete é colocado em *freezers* e comercializado.

Na sede da empresa, a vida de cada litro de sorvete – desde os ingredientes até a venda – é acompanhada de perto. Depois que um litro é rotulado e enviado, a Ben & Jerry's armazena seu número de rastreamento em um *data warehouse* da Oracle e posteriormente analisa os dados. Usando *software* de BA, a equipe de vendas consegue verificar se o Chocolate Chip *Cookie Dough* está ganhando terreno sobre o Cherry Garcia no cobiçado posto de líder de vendas. O departamento de *marketing* confere se as

promoções e propagandas da empresa estão gerando um aumento nas vendas. O pessoal de finanças usa os números de rastreamento em suas análises para mostrar a renda gerada por cada tipo de sorvete. Desde que a empresa começou a usar o *software*, o departamento de contabilidade reduziu drasticamente o tempo que leva para encerrar o balanço mensal. Também, provavelmente o mais importante para uma empresa focada na fidelidade do cliente, a equipe de assuntos do consumidor faz a correspondência de cada litro com as centenas de ligações e *e-mails* recebidos toda semana para verificar se há alguma reclamação.

Fontes: compilado de J. Schlosser. "Looking for Intelligence in Ice-Cream" *Fortune*, March 17, 2003; and Ben & Jerry's Homemade Inc./Information Systems, essaypage.com/cgi-bin/query?mss=essaypage&stq=10&q=Ben%20and%20Jerry%27s (acessado em julho de 2006).

Aplicações mais avançadas de BA incluem atividades como modelagem financeira, orçamentos, alocação de recursos e inteligência competitiva. Os sistemas avançados de BA envolvem componentes como modelos de decisão, análise do desempenho dos negócios, métricas, perfis de dados, ferramentas de reengenharia e muito mais. (Para ver detalhes, visite dmreview.com.) Por fim, a BA pode ser realizada em tempo real (consulte Bonde and Kuckuk: 2004 e IT BI da Microsoft em microsoft.com/sql/solutions/bi/overview.msp).

As ferramentas e técnicas de BA

A BA emprega um grande número de ferramentas e técnicas de análise. Podemos dividi-las em três grandes categorias, conforme apresentado na Figura 3.1.

Como mostra a figura, a primeira categoria é a *descoberta de informações e conhecimento*. Algumas de suas atividades são discutidas neste capítulo. *Data, text e web mining* são descritos no Capítulo 4. A segunda categoria é o *suporte à decisão e sistemas inteligentes*, descritos brevemente aqui. A categoria final, *visualização*, tem discussões aqui e no Capítulo 5.

Os fornecedores classificam as ferramentas de BA de várias formas diferentes. Apresentaremos duas delas a seguir.

Classificação das ferramentas de BA pela MicroStrategy: os cinco estilos de BI

A MicroStrategy divide seus produtos em cinco categorias (chamadas de *estilos*; veja os detalhes em Microstrategy: 2006). Eis os cinco estilos:

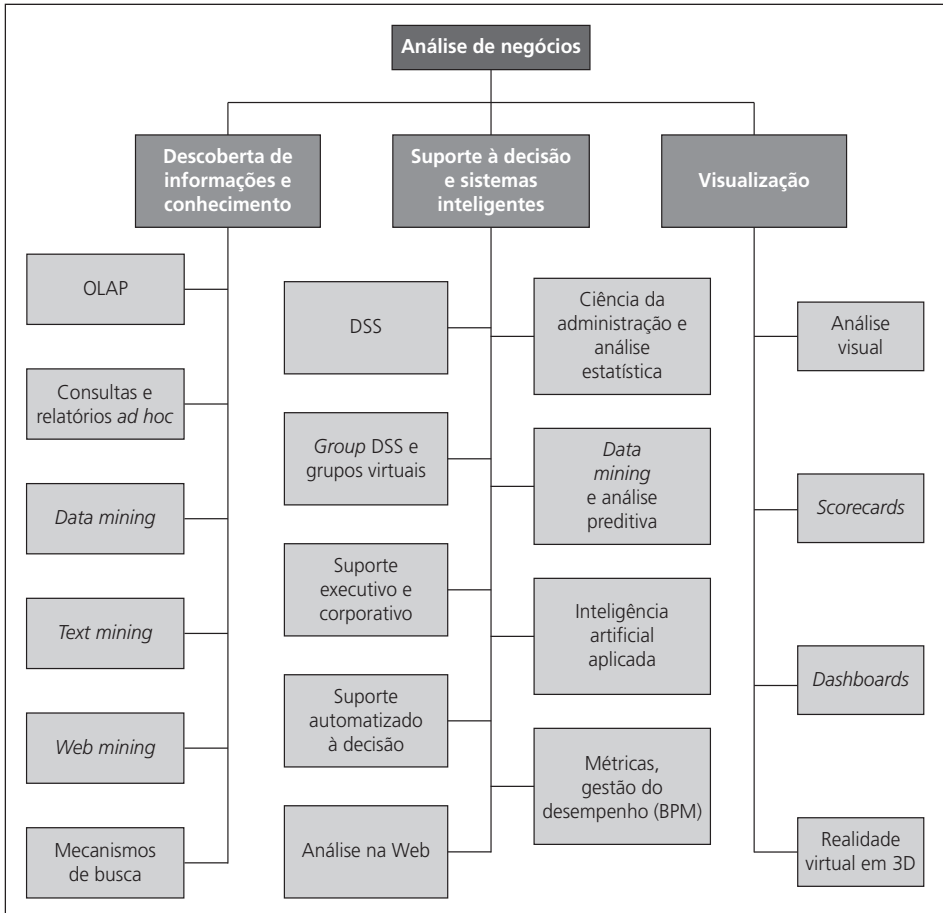


Figura 3.1 Categorias de análise de negócios.

1. **Relatórios empresariais.** Os produtos de relatórios empresariais são usados para gerar relatórios estáticos altamente formatados, para ampla distribuição a muitas pessoas. São formatos de relatório *pixel-perfect* para relatórios operacionais e *dashboards*.
2. **Análise de cubos.** Ferramentas de BI baseadas em cubos são usadas para oferecer recursos analíticos com visões parciais multidimensionais simples de OLAP aos gerentes, em um ambiente limitado.
3. **Consulta e análise ad hoc.** Ferramentas OLAP relacionais são usadas para permitir que os usuários avançados façam uma consulta no banco de dados buscando qualquer resposta, executem com visões parciais em todo o banco e realizem *drill down* até o nível mais detalhado de informações transacionais. Esta consulta investigativa é direcionada aos exploradores de informações e aos usuários avançados.
4. **Análise estatística e data mining.** São usadas ferramentas estatísticas, matemáticas e de *data mining* para executar análise preditiva ou para descobrir a correlação de causa e efeito entre duas métricas. Análises e previsões financeiras também são executadas.
5. **Entrega de relatórios e alertas.** Os mecanismos de distribuição de relatórios são usados de maneira proativa para enviar relatórios completos ou alertas para grandes populações de usuários (internos e externos), com base em assinaturas, programações ou eventos limitados nos bancos de dados.

O principal produto da MicroStrategy – atualmente, o MicroStrategy 8 – é construído segundo essa classificação (veja a Figura 3.2). Observe que a arquitetura permite uma única estrutura *plug-and-play* unificada, para que os usuários encontrem todas as ferramentas necessárias integradas em um só lugar.

Outros fornecedores de BI oferecem classificações semelhantes a seus produtos. Uma exceção digna de nota é a SAP AG.

Classificação da gestão estratégica de empresas pela SAP: operacional, gerencial e estratégico

A SAP AG (sap.com) é a maior fornecedora mundial de produtos de *software* corporativo, como ferramentas de planejamento de recursos empresariais (ERP). A empresa desenvolve seus produtos de acordo com três níveis de suporte:

1. *Operacional*. O ERP (SAP R/3) suporta principalmente o processamento de transações no nível operacional.

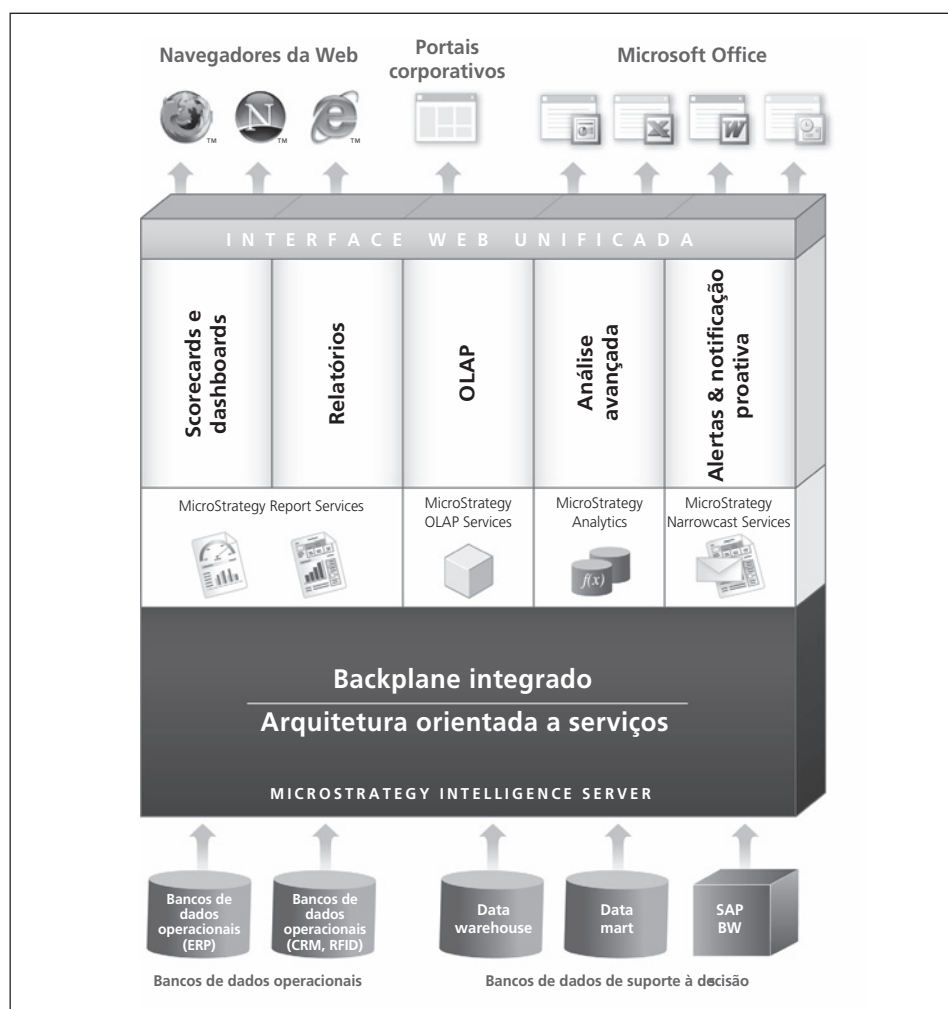


Figura 3.2 Arquitetura do MicroStrategy 8.

Fonte: microstrategy.com/Software/index.asp.

- 2. *Gerencial*. Em termos gerenciais, os gerentes de nível médio usam o SAP/R3 para acessar todos os relatórios, organizados por áreas funcionais (como *marketing*, finanças). Eles podem realizar consultas e *drill down*.
- 3. *Estratégico*. No nível estratégico, a empresa oferece produtos sob o título de SAP SEM (Strategic Enterprise Management, *gestão estratégica de empresas*, em português), que incluem BA (SAP SEM/BA).

Para mais descrições dos níveis da SAP, consulte Thompson e Jakovljevic (2005), Meier et al. (2005) e **sap.com**.

Dividimos este capítulo em seções parecidas com a classificação da MicroStrategy. Antes de apresentarmos os detalhes, examinaremos algumas atividades populares de BA, que em sua maioria são herança dos recursos de *sistema de informações executivas (EIS)*.

Sistemas de informações executivas e de suporte executivo

Muitas atividades de BI tiveram origem em duas ferramentas:

- *Sistemas de informações executivas (EIS)*. Um EIS é um sistema baseado em computadores que serve as necessidades dos altos executivos por informação. Ele oferece um acesso rápido a informações pontuais e relevantes a fim de melhorar o crescimento e o aprendizado gerencial, e também de auxiliar no monitoramento do desempenho de uma organização através do acesso direto aos relatórios de gestão. Um EIS é muito fácil de usar, é auxiliado por gráficos e oferece os recursos de *relatórios de exceção* (ou seja, relatórios apenas sobre os resultados que diferem de um padrão estabelecido) e *drill down* (isto é, a investigação cada vez mais detalhada das informações). Um EIS pode ainda ser conectado a serviços de informações *online* e *e-mail*.
- *Sistemas de suporte executivo (SSE)*. Um SSE é um abrangente sistema de suporte que vai além do EIS e inclui suporte à análise, comunicações, automação do escritório e suporte à inteligência.

O EIS e o SSE variam em suas capacidades e benefícios. As capacidades em comum a muitos EIS/SSE encontram-se resumidas na Tabela 3.1.

Há detalhes de algumas destas capacidades no Arquivo *online* W3.1. As listadas na Tabela 3.1 são comuns nas várias ferramentas de BA apresentadas nas próximas seções. Uma destas capacidades, o fator crítico de sucesso (FCS), é medida pelos indicadores-chave de desempenho (KPI)

Tabela 3.1 Capacidades de EIS/SSE

Capacidade	Descrição
<i>Drill down</i>	Capacidade de chegar a mais detalhes em um ou vários níveis. Pode ser feito por meio de uma série de menus ou por consultas diretas (usando agentes inteligentes e processamento de linguagem natural).
Fatores críticos de sucesso (FCS)	Os fatores mais críticos ao sucesso da empresa. Podem ser organizacionais, departamentais, do setor etc.
Indicadores-chave de desempenho (KPI)	As medidas específicas de cada FCS.
Relatórios de status	Os dados mais recentes sobre KPIs ou alguma outra métrica; nas condições ideais, são fornecidos em tempo real.
Análise de tendências	Tendências de KPIs ou métricas a curto, médio e longo prazo, projetadas com o uso de métodos de previsão.
Análise <i>ad hoc</i>	Análise realizada a qualquer momento e com qualquer fator ou relação desejado.
Relatórios de exceção	Uso de relatórios que destacam os desvios que superam determinados limites. Os relatórios podem incluir apenas desvios.
Visões parciais de cubos	Reorganização dos dados para que sejam visualizados de diferentes perspectivas.

e apresentada no Capítulo 5. Muitas ferramentas de BA aparecem sob o título de *processamento analítico online (OLAP)*, as quais descreveremos a seguir.

Questões de revisão da Seção 3.2

1. Defina a BA e explique seu valor estratégico.
2. Relacione a BA ao *data warehousing*.
3. Liste e descreva as três categorias de ferramentas exibidas na Figura 3.1.
4. Liste os cinco estilos de BI da MicroStrategy.
5. Descreva a categorização de ferramentas de BI elaborada pela SAP.
6. Defina *EIS* e *SSE*.
7. Liste as principais capacidades de EIS/SSE (consulte a Tabela 3.1).

3.3

PROCESSAMENTO ANALÍTICO ONLINE (OLAP)

O termo **processamento analítico online (OLAP)** se refere a uma variedade de atividades normalmente executadas por usuários finais em sistemas *online*. Não há um consenso sobre quais atividades são consideradas OLAP. Normalmente, OLAP inclui atividades como geração e resposta de consultas, solicitação de relatórios e gráficos *ad hoc* e execução dos mesmos, realização de análises estatísticas tradicionais ou modernas e construção de apresentações visuais. Muitas pessoas também pensam em análise e apresentações multidimensionais, EIS/SSE e *data mining* como atividades de OLAP. Basicamente os produtos de OLAP oferecem recursos de modelagem, análise e visualização de grandes conjuntos de dados, ou para sistemas de gerenciamento de banco de dados (SGBD) ou, mais freqüentemente, para sistemas de *data warehouse*. Os produtos oferecem também uma visão conceitual mutidimensional dos dados.

OLAP versus OLTP

Durante muitos anos, a TI concentrou-se em construir sistemas essenciais que permitissem principalmente o *processamento de transações* corporativas. Tais sistemas deveriam ser praticamente tolerantes a falhas e oferecer execução eficiente e resposta rápida. O *processamento de transações online (OLTP)* ofereceu uma solução eficaz, que gira em torno de atividades repetitivas e de rotina, usando um ambiente de *banco de dados relacional* distribuído. Os últimos desdobramentos nesta área são o uso de *softwares* de ERP e de gestão da cadeia de fornecimento (SCM) para tarefas de processamento de transações, aplicações de gerenciamento de relacionamento com o cliente (CRM) e integração com tecnologias baseadas na Web e *intranets*. Foram criadas muitas ferramentas para desenvolver aplicações de OLTP. O Informix Dynamic Server (www-306.ibm.com/software/data/informix/ids) é um exemplo de uma boa ferramenta desse tipo.

Tanto as aplicações de OLTP quanto de sistemas de suporte à gerência (MSS) constantemente necessitam de acesso aos dados. Infelizmente, tentar servir os dois tipos de solicitação pode ser uma tarefa problemática (consulte Gray and Watson: 1998). Portanto, algumas empresas escolhem separar os sistemas de informações em tipos OLTP e tipos OLAP.

O OLTP é voltado para o processamento de transações repetitivas em grandes quantidades e manipulações simples. O OLAP envolve o exame de muitos itens de dados (constantemente milhares ou até milhões) em *relacionamentos complexos*. Além de responder às consultas dos usuários, o OLAP consegue analisar esses relacionamentos e buscar padrões, tendências e exceções. Em outras palavras, o OLAP é um método direto de suporte à decisão.

Uma consulta de OLAP pode acessar um banco de dados com *gigabytes* ou *terabytes* de informações sobre vários anos de vendas a fim de encontrar todas as vendas de produtos em cada região para cada tipo de produto. Após revisar os resultados, um analista pode refinar ainda mais a pesquisa com o objetivo de encontrar o volume para cada canal de vendas dentro de uma região, ou dentro de

determinadas classificações do produto. Uma última etapa pode envolver o analista executando comparações entre anos ou trimestres para cada canal de vendas. O processo inteiro deve ser executado *online*, com um tempo rápido de resposta para que o processo de análise ocorra sem interferências.

O OLAP também se diferencia do OLTP (e do *data mining*) no sentido de que os usuários podem fazer perguntas específicas e abertas. Os usuários, normalmente analistas, *guiam* o OLAP, ao passo que o *data mining* busca relações usando apenas alguma orientação do analista. O trabalho com o *data warehouse* (ou com *data marts* ou um banco de dados multidimensional) e com um conjunto de ferramentas OLAP facilita o processamento analítico *online*. Essas ferramentas podem ser planilhas, ferramentas de consulta, de *data mining*, de visualização de dados e semelhantes. Para uma lista de ferramentas OLAP, consulte Alexander (2003), **baseline.com** e revisões periódicas nas seções de *software* de *PCWeek*, *DM Review*, *Intelligent Enterprise* e *Software Review*. Os principais fornecedores dessas ferramentas incluem Business Objects, Computer Associates, Cognos Inc., Geas, Hyperion Software Corp., Informatica Corp., Information Builders, IBM, Intersolve, Microsoft, MicroStrategy Corp., Oracle, SAS Institute, Inc., SPSS e Temtec.

Tipos de OLAP

Os principais tipos de OLAP são:

- *OLAP multidimensional (MOLAP)*. Quando o OLAP é implementado através de um banco de dados (ou data store) multidimensional especializado, ele é chamado de **OLAP multidimensional (MOLAP)**, porque resume transações em visões multidimensionais com antecedência (consulte a Seção 3.5). Os dados são organizados em uma *estrutura de cubos* que o usuário pode girar, o que é adequado principalmente a resumos financeiros. Com o MOLAP, as consultas são mais rápidas pois a consolidação já foi feita.
- *OLAP relacional (ROLAP)*. Quando um banco de dados OLAP é implementado sobre um banco de dados relacional existente, ele é chamado de **OLAP relacional (ROLAP)**. As ferramentas OLAP relacional extraem dados de bancos de dados relacionais. Ao usar declarações de SQL complexas com relação a tabelas relacionais, o ROLAP pode também criar visões multidimensionais dinamicamente. O ROLAP tende a ser usado em dados que apresentam um grande número de atributos, onde não possam ser colocados facilmente em uma estrutura de cubos. Por exemplo, os dados do cliente com diversos campos descritivos, ao contrário dos dados financeiros, são normalmente candidatos a ROLAP.
- *Database OLAP e Web OLAP (DOLAP e WOLAP)*. O *database OLAP* se refere a um sistema de gerenciamento de banco de dados relacional (SGBDR) projetado para hospedar estruturas e executar cálculos de OLAP. O *Web OLAP* se refere aos dados de OLAP acessíveis de um navegador da Web.
- *Desktop OLAP*. O desktop OLAP envolve ferramentas OLAP simples e baratas, que executam análise local multidimensional e apresentação de dados baixados de bancos de dados relacionais ou multidimensionais para as máquinas do cliente. Versões da Web constantemente movem o processamento de *desktop* para um servidor intermediário, o que aumenta a escalabilidade; no entanto, na melhor das hipóteses, a funcionalidade é comparável à da versão de *desktop*.

Exemplos de aplicações OLAP bem-sucedidas

A ING Antai Insurance Co. (Taiwan) usa o OLAP Server da IBM para analisar grandes quantidades de dados, a fim de detectar pedidos fraudulentos e acelerar o processamento dos pedidos. Agora, ela leva somente cerca de dois dias para analisar os dados, o que antes levava algumas semanas. Conforme as ferramentas e o *hardware* são aperfeiçoados, os pedidos são analisados com mais rapidez – ou até mesmo instantaneamente. O custo com o processamento de pedidos teve uma enorme redução. Outro exemplo é o TCF Bank, que faz uso do OLAP para oferecer as

informações diretamente a quem precisa. O banco agora tem uma compreensão mais exata de sua base de clientes, para poder direcionar melhor seus empreendimentos de *marketing*.

De acordo com a Cognos.com (2002), a Allied Building Products Corp. aumentou sua companhia de distribuição de materiais de construção ampliando sua fatia de mercado e clientes. A Allied implementou o Cognos Finance, uma solução de BI da Cognos, em menos de 90 dias. O resultado disso foi que a Allied conseguiu padronizar seus dados por toda a empresa e automatizar o processamento, podendo oferecer uma visão única e coordenada do desempenho financeiro. O Cognos Finance permite que a empresa reduza o trabalho manual na produção de relatórios e, ao mesmo tempo, dá acesso a orçamentos, previsões e mercadorias reais nas operações de todas as suas filiais. Isso permitiu uma visão integrada das informações e acelerou a geração de relatórios financeiros acurados.

A Tabela 3.2 proporciona mais exemplos de aplicações OLAP em diversas organizações.

Tabela 3.2 Aplicações OLAP representativas

<i>Empresa</i>	<i>Descrição</i>
British Petroleum Exploration and Production (bp.com)	A BP usa um sistema de planejamento mundial e em grande escala, agora em sua segunda geração baseada em OLAP, que substituiu uma implementação mais complexa, baseada em planilhas.
Bromford Iron & Steel Co. Ltd. (bromfordsteels.co.uk)	Este é um exemplo raro de aplicação OLAP usada para analisar dados detalhados de produção para um pequeno fabricante de aço. O objetivo é aumentar a eficiência, reduzir o consumo de energia, refinar a estratégia de produtos e melhorar o desempenho financeiro.
Blue Rhino (bluerhino.com)	Mesmo empresas menores com mínimos recursos internos de TI podem aproveitar ao máximo as aplicações OLAP flexíveis por toda a organização.
Time Warner (timewarner.com)	A Time Warner oferece suporte aos usuários em três continentes, com um planejamento estratégico de mercado e sistema de análise que usa um servidor de banco de dados OLAP.
The World Bank (worldbank.org)	The World Bank usa ferramentas OLAP para executar análises estatísticas complexas em uma massa de dados econômicos mundiais.
British Airways (ba.com)	A British Airways reduziu seus custos com processamento e obteve melhores recursos de análise usando bancos de dados OLAP juntamente com um novo sistema contábil.
Deluxe Corp. (deluxe.com)	As aplicações de planejamento e análise economizaram US\$ 10 milhões; também ofereceram recursos mais exatos de previsão para a empresa número um do mundo em impressão de cheques.
Dun & Bradstreet (dnb.com)	A tecnologia de OLAP propicia uma conexão vital entre o resultado da análise dos dados corporativos dos clientes e as informações detalhadas subjacentes, um ingrediente essencial na venda de informações comerciais.
GlaxoSmithKline (gsk.com)	Bancos de dados OLAP são usados para a geração de relatórios financeiros internacionais a partir de um sistema de consolidação em <i>mainframe</i> .
Subaru of America (subaru.com)	Uma ferramenta OLAP ajudou a empresa a ter mais eficiência na alocação de estoques para revendedores franqueados, com o propósito de minimizar as retenções de estoque e maximizar as vendas e o atendimento ao cliente.
Sun Microsystems (sun.com)	Ao pôr em prática aquilo que defende sobre computação centrada em rede, a Sun usou uma ferramenta OLAP em rede para implementar uma nova aplicação de planejamento de negócios. A aplicação é totalmente baseada na Web e usa agentes inteligentes para criar <i>business intelligence</i> .
Ford Motor Co. (ford.com)	Para responder às consultas de 5.500 concessionários, centenas de locadoras de automóveis e outros, a Ford equipou os funcionários de mais de 300 centrais de atendimento com acesso a informações desenvolvidas por BI.

Fontes: compilado de *The OLAP Report*. **olapreport.com** (acessado em janeiro de 2006) e de vários sites de fornecedores de BI.

Ferramentas e fornecedores de OLAP

O uso de SQL e outras ferramentas convencionais de acesso e análise de dados é útil, mas não suficiente para o OLAP. Em OLAP é usada uma classe especial de ferramentas, incluindo *front-ends* de BA, de acesso aos dados, de banco de dados e sistemas de acesso a informações visuais. Além disso, os métodos de OLAP superam em muito as planilhas em termos de potência e resultados: sua intenção é dar *autonomia* aos usuários. Para maiores detalhes, consulte Perry e Post (2007).

Características das ferramentas OLAP

As ferramentas OLAP têm características que as diferenciam das ferramentas de SIG cujo propósito é suportar aplicações tradicionais de relatórios de OLTP. E. F. Codd et. al. (1993) definiram sucintamente as características das ferramentas OLAP nas 12 regras resumidas na Tabela 3.3. As 12 regras são usadas para padronizar a modelagem de dados multidimensionais e definem quatro tipos de processamento executados pelos analistas em uma organização:

1. A *análise categórica* é uma análise estática baseada em dados históricos. Ela se vale da premissa que o desempenho passado é um indicador do futuro. Esta é a análise básica suportada por bancos de dados OLTP baseados em transação.
2. A *análise exegética* também toma como base os dados históricos e acrescenta a capacidade de análise *drill down*. A *análise drill down* é a capacidade de consultar os dados mais a fundo para determinar os dados detalhados usados para determinar um valor derivado.
3. A *análise contemplativa* permite que um usuário altere um único valor a fim de determinar seu impacto.
4. A *análise formalista* permite alterações a múltiplas variáveis.

Ferramentas OLAP representativas

Os fornecedores no campo de BI estão usando de artifícios para dar aos usuários finais a capacidade de customizar aplicações analíticas e satisfazer as necessidades do negócio em evolução. As aplicações incluem, por exemplo, a plataforma de análise DecisionSite da Spotfire (**spotfire**).

Tabela 3.3 Regras de avaliação de produtos OLAP: as 12 regras de Codd para o OLAP

1. Visão conceitual mutidimensional para formular consultas.
2. Transparência ao usuário.
3. Fácil acessibilidade: acesso *online* e *batch*.
4. Desempenho consistente na geração de relatórios.
5. Arquitetura cliente/servidor: o uso de recursos distribuídos.
6. Dimensionalidade genérica.
7. Manipulação dinâmica da matriz esparsa.
8. Suporte multiusuário, ao invés de suporte a um único usuário.
9. Operações irrestritas com dimensões cruzadas.
10. Manipulação intuitiva dos dados.
11. Relatórios flexíveis.
12. Dimensões e níveis de agregação ilimitados.

Fontes: adaptado de E.F. Codd, S.B. Codd, and C.T. Salley, *Providing OLAP to User-Analysts: An ITn andate*, dev.hyperion.com/resource_library/white_papers/providing_olap_to_user_analysts.pdf (acessado em fevereiro de 2006); and M. Meier, W. Sigzig, and P. Mertens, *Enterprise Management with SAP SEM/Business Analytics*, 2nd ed., Berlin: Springer, 2005.

com), a Enterprise BI Suite da Business Objects (**businessobjects.com**) e a QlikView da QlikTech (**qliktech.com**). Veja os detalhes em Havenstein (2003b).

O Temtec Executive Viewer fornece todos os principais recursos de OLAP, incluindo visões multidimensionais dos dados, expansão e recolhimento de dimensões, seleção dinâmica de colunas, cálculos automáticos (somas etc.), produção automática de gráficos e tabelas, mapas físicos para exibir dados (consulte a Seção 3.7) e *drill down* e *roll up* instantâneos.

Para saber mais sobre OLAP, consulte Paris Technologies (**olap.com**) e Havenstein (2003a). Para outra aplicação concreta de OLAP, leia o Caso de aplicação 3.2.

Um dos principais benefícios do OLAP para os usuários finais (conforme relatado por **temtec.com**) foi a eliminação da necessidade de escrever relatórios e resumir dados manualmente (ou semimanualmente). As principais atividades de OLAP – relatórios, consultas e análise – são descritas mais detalhadamente nas próximas seções.

Questões de revisão da Seção 3.3

1. Defina *OLAP* e compare ao OLTP.
2. Descreva as principais características e benefícios do OLAP.
3. Descreva os tipos de OLAP.
4. Liste e descreva brevemente as principais ferramentas de OLAP.

CASO DE APLICAÇÃO 3.2

TCF Financial Corp.: conduzindo OLAP, relatórios e *data mining*

Um dos maiores bancos regionais do centro-oeste dos Estados Unidos, o TCF Bank (**tcfbank.com**) tem mais de 400 filiais em seis estados e atende clientes de todos os níveis de renda. O TCF Bank também opera o quarto maior sistema do país de filiais de banco em supermercados. A empresa se concentra em ser um local único de parada para os clientes: é um dos poucos bancos dos EUA a abrir 12 horas por dia, 7 dias por semana, inclusive em feriados.

Os usuários dos principais grupos do banco (como varejo, empréstimos, hipotecas e corretagem) descobriram que os relatórios de TI não supriam suas necessidades de suporte à decisão. Assim, eles tiveram de desenvolver processos personalizados para baixar arquivos dos dados operacionais brutos, depois carregar os dados em planilhas para analisá-los melhor. O tempo necessário para criar um relatório gráfico padrão era até um mês; seis semanas era o que se levava para gerar uma lista de *marketing* de clientes.

O departamento de gerenciamento de informações do TCF precisava criar um processo melhor, que permitisse aos usuários obter a percepção dos clientes para revelar oportunidades e oferecer-lhes novos serviços de forma eficaz. O banco, desta for-

ma, adotou o PowerCenter e o PowerAnalyzer da Informatica Corp. em meados de 2002. Os recursos de assistente de criação de relatórios, relatórios baseados em métricas e *drill down* do caminho de análise do PowerAnalyzer foram fatores importantes na decisão pela adoção, visto que suas funções eram fáceis de usar. Também foram desenvolvidos vários relatórios iniciais de indicadores-chave para *dashboards* de usuários. Em uma semana, 550 funcionários do setor de empréstimos e executivos estavam usando este e outros relatórios diariamente.

Com o novo sistema OLAP, que inclui uma aplicação de vendas cruzadas, o TCF consegue identificar classes de clientes a quem abordar com produtos e serviços específicos correspondentes. Isso é particularmente crucial na identificação das necessidades de clientes novos. Além do mais, são gerados relatórios e é fornecido OLAP para facilitar análises adicionais.

Fontes: adaptado T. Ledman, “TCF Bank”, *What Works: Best Practices in Business Intelligence and Data Warehousing*, Vol. 15, The Data Warehousing Institute, Chatsworth, CA, June 2003, p. 10; em **securityindustry.com**, June 20, 2005; and **informatica.com** (acessado em fevereiro de 2006).

RELATÓRIOS E CONSULTAS

As atividades mais antigas de OLAP e BI estão usando relatórios e consultas. Eles freqüentemente estão correlacionados porque, em muitos casos, o resultado de uma consulta é um relatório, e um relatório pode designar uma consulta. O OLAP permite ao usuário produzir facilmente seus próprios relatórios/panoramas e analisar tendências e desempenho diariamente.

Relatórios

De acordo com as regras de Codd, a geração de relatórios OLAP deve ser uniforme, flexível e ajustável. Dois tipos de relatório são diferenciados: rotina e *ad hoc*.

Relatórios de rotina

Os *relatórios de rotina* são gerados automaticamente e distribuídos periodicamente aos assinantes em listas de discussão. Alguns exemplos são números de vendas semanais, unidades produzidas por dia e por semana, e horas trabalhadas mensalmente. Segue um exemplo de como um relatório é usado em BI: uma gerente de loja recebe relatórios de desempenho da loja gerados semanalmente por um *software* de BI. Após uma análise de um determinado relatório semanal sobre as vendas da loja, a gerente observa que as vendas de periféricos para computador caíram significativamente desde as semanas anteriores. Ela clica no seu relatório e imediatamente vai para outro relatório da empresa, que lhe mostra que os três discos rígidos mais vendidos surpreendentemente estão rendendo menos que o esperado. Agora, a gerente precisa investigar o motivo. Um novo detalhamento por dia individual deve revelar que o mau tempo causou o problema.

Os relatórios podem ser gerados diretamente de dados operacionais (p.ex., ERP, sistemas de pontos de venda [PDV]) e/ou de um *data warehouse*.

Relatórios ad hoc (ou sob demanda)

Os relatórios *ad hoc* são criados para um usuário específico sempre que necessário. Esses relatórios podem ser semelhantes aos de rotina, mas com intervalos de tempo diferentes ou apenas para um subconjunto de dados. Um exemplo seria fornecer uma lista de todos os clientes que adquiriram produtos de uma empresa por mais de US\$5.000 por produto, durante janeiro de 2006.

Suporte multilíngüe

Muitos fornecedores oferecem tradução de relatório para diversos idiomas (p.ex., a Microsoft para 12 idiomas). Este suporte inclui toda a ajuda na interface (p.ex., barras de menus, grupos de caracteres, conversão e formatação de moeda, atributos do negócio).

Exemplos de produtos do fornecedor para relatório

Todos os fornecedores de *software* de BI oferecem recursos para criação, acesso e manutenção de relatórios. Exemplos e demos de relatórios estão disponíveis nos *sites* da maioria dos fornecedores de BI. Os exemplos típicos são:

- *Crystal Reports* da *Business Objects* é um conjunto de ferramentas que ajuda a criar relatórios flexíveis e cheios de recursos de maneira rápida e os integra às aplicações do Windows e da Web.
- *MicroStrategy* fornece ferramentas de criação e monitoramento de relatório para relatórios operacionais e de produção, como faturas e extratos, relatórios de lucro e perda, e relatórios de desempenho. Estão disponíveis relatórios personalizados, assim como suporte multilíngüe (12 idiomas). Os relatórios possuem resolução de impressão, tela para gráficos e tabelas de qualidade.

- Gerador de relatório *Cognos 8 Business Intelligence* inclui uma lista completa de tipos de relatório automático, ajustáveis a qualquer fonte de dados. Também fornece suporte multilíngüe.
- *Hyperion* fornece um espectro completo de recursos para gerenciamento de relatórios que combina tanto informações operacionais como financeiras. É possível personalizar relatórios em tela em grande quantidade para publicação para dezenas de milhares de clientes através da Web com uma taxa de transferência excelente. Um módulo para geração de relatório financeiro especializado, com formatação e inteligência financeira predefinidas, permite reunir e publicar de maneira rápida registros de relatórios de qualidade da produção para o controle do gerenciamento e arquivamentos regulares. Para relatórios e perguntas *ad hoc*, os usuários de negócios também podem criar seus próprios relatórios interativos para monitoramento do seu desempenho e reconhecimento de tendências. A *Hyperion* também oferece geração de relatório interativo.
- A *Microsoft* incluiu no seu Report Builder um recurso de fácil utilização que permite a criação ou modificação de relatórios sem a necessidade de escrever consultas SQL (veja Swoyer, 2005).

Tipos representativos de relatórios

O *software* de BI pode ser usado para produzir dúzias de relatórios em todas as áreas funcionais. Para exemplos representativos, consulte a Tabela 3.4.

Indicadores e dashboards

Indicadores e *dashboards* são considerados ferramentas de relatório extensivas porque fornecem panoramas gráficos e tabulares de diversos relatórios, incluindo comparações de métricas. Eles são apresentados no Capítulo 5. De forma semelhante, a visualização de dados (consulte Seção 3.7) inclui a apresentação visual de relatórios.

Tabela 3.4 Relatórios típicos produzidos por OLAP

Área	Descrição
Finanças/contabilidade	Relatório e análise de contas a receber e a pagar, relatório e análise de balancete, análise do fluxo de caixa, previsão e orçamento financeiros, análise da demonstração de resultados/lucro e perda, gerenciamento de risco
Recursos humanos	Relatório dos benefícios da administração, relatório e análise do plano de carreira, relatório de conformidade, relatório e análise de compensação, relatório de gastos com funcionários, análise de tendência da força de trabalho, relatório de gerenciamento de tempo e trabalho
Previsão de vendas e marketing	Análise de canal, análise da concorrência, relatório de desempenho da equipe de vendas, relatório e análise do fluxo de vendas
Gerenciamento de relacionamento com o cliente	Análise de campanha; análise de aquisição, retenção e desgaste do cliente; análise de venda cruzada para o cliente; análise da lucratividade do cliente; indicadores do cliente; análise de segmentação do cliente; análise do atendimento ao cliente; análise de fidelidade
Vendas de produto	Análise de desempenho da categoria/gerenciamento da categoria, análise de descontos e promoções, análise de fraude e prevenção de perda, análise do estoque, análise de <i>merchandising</i> , indicadores do produto
Operação e gestão da cadeia de fornecimento	Análise da central de atendimento, planejamento da produção e programação da manufatura, otimização da rede, análise de cumprimento do pedido, controle de qualidade, gerenciamento normativo, indicadores do fornecedor e análise de fornecimento estratégico, planejamento de oferta e demanda
Análise de site	Análise do comércio eletrônico, análise de navegação na Web, análise de tráfego na Web, análise de visitantes da Web

Alerta e entrega de relatório

A grande maioria dos fornecedores de *software* de BI oferece *alerta e entrega de relatório* para distribuir proativamente um grande número de relatórios e avisa um número potencialmente grande de usuários (tanto funcionários internos como externos da empresa). Por exemplo, os produtos de *software* podem distribuir centralmente *e-mails* para uma extensa população de usuários, com anexos e em uma base programada. Além disso, alguns fornecedores oferecem recursos adicionais, como, por exemplo, os seguintes, oferecidos pela MicroStrategy:

- *Distribuição de relatório através de qualquer ponto de contato.* Uma empresa pode usar *e-mail*, impressoras, servidores de arquivo e *sites* com contêineres HTML, Excel, RTF, PDF e ZIP para transportar o conteúdo do relatório.
- *Auto-assinatura e distribuição baseada em administrador.* Uma empresa pode combinar, de forma centralizada, distribuições de relatório compulsórias com necessidades de informação direcionada a cada indivíduo.
- *Entrega sob demanda, programada ou na ocorrência de evento.* Uma empresa pode ativar a distribuição de relatório por quaisquer meios que sejam mais adequados ao usuário.
- *Personalização automática de conteúdo.* Uma empresa pode fazer um relatório mais relevante e seguro para todos os usuários. Relatórios personalizados economizam o tempo dos usuários ao mostrar somente informações e dados solicitados.

Para obter mais detalhes, consulte MicroStrategy (2006).

Relatórios e consultas

Consultas ad hoc

Qualquer consulta que não pode ser determinada antes de ser feita será considerada uma **consulta ad hoc**. O usuário pode decidir inserir tal consulta após receber um relatório. Consultas *ad hoc* permitem aos usuários solicitar, a partir do computador, informações que não estão disponíveis nos relatórios periódicos e também gerar consultas novas ou modificar as antigas, com flexibilidade significativa em termos de conteúdo, estrutura e avaliações. Essas respostas são necessárias para agilizar ou facilitar a tomada de decisão. O sistema deve ser inteligente o suficiente para compreender o que o usuário deseja. Sistemas simples de consulta *ad hoc* muitas vezes são baseados em menus. Sistemas mais inteligentes usam SQL e abordagens de consulta através de exemplo. Sistemas muito inteligentes estão baseados em entendimento de linguagem natural e alguns conseguem se comunicar com os usuários usando reconhecimento de voz. As consultas podem ser feitas em dados estáticos ou dinâmicos (ou seja, em tempo real). (Ainda neste capítulo, abordaremos o uso de ferramentas da Web para facilitar as consultas.) Por fim, a BusinessObjects fornece a ferramenta Intelligent Question que leva o usuário a fazer as perguntas certas.

SQL para consulta

A **Linguagem de Consulta Estruturada (SQL)** é uma linguagem padrão para acesso e manipulação de dados em DBMS relacional. É uma linguagem semelhante ao inglês que consiste em diversas camadas de complexidade e capacidade crescentes. A SQL é usada para acesso *online* a bancos de dados, operações DBMS, a partir de programas, e funções administrativas em bancos de dados. Também é usada em funções de acesso e manipulação de dados em alguns dos principais produtos de *software* DBMS (p.ex., Oracle 9i e 10i, DB2 da IBM, Microsoft SQL Server 2005). Para obter mais detalhes, consulte Perry e Post (2007).

Como a SQL é não-processual e razoavelmente fácil de usar, muitos usuários finais conseguem usá-la para criar suas próprias consultas e operações de banco de dados. A SQL pode ser usada para programas escritos em qualquer linguagem padrão de programação; por isso, facilita a integração de *softwares*. As consultas também podem ser realizadas em linguagens naturais.

Freqüentemente, os sistemas de consulta são combinados com sistemas de relatório. Para um exemplo dessa combinação em uma locadora de vídeo, consulte Amato-McCoy (2003).

Análise de resultados do relatório

Em muitos casos, os dados fornecidos pelos relatórios exigem uma investigação adicional. Essa investigação precisa ser realizada razoavelmente rápido – algumas vezes imediatamente após a visualização dos dados – e deve ser feita pelos usuários finais a um baixo custo. Por essa razão, os fornecedores oferecem ferramentas de relatório que permitem uma investigação adicional. Por exemplo, um detalhamento rápido pode oferecer uma explicação para vendas perdidas, ou uma análise de tendência pode gerar um alerta. Muitas ferramentas são fornecidas com produtos OLAP. Esse tipo de análise pode ser facilitado com a multidimensionalidade, que será abordada a seguir.

Questões de revisão da Seção 3.4

1. Defina *relatório de rotina* e *relatório ad hoc*.
2. Liste as categorias de relatório mais importantes (segundo a MicroStrategy).
3. Que tipo de relatório está incluído nos *kits* dos principais fornecedores?
4. Explique por que os indicadores e *dashboards* são vistos como categorias especiais de relatório.
5. Defina *entrega e alerta de relatório*.
6. Defina *consulta ad hoc*. Como uma consulta *ad hoc* pode ser gerada?

3.5

MULTIDIMENSIONALIDADE

Dados brutos e concisos podem ser organizados de maneiras diferentes para análise e apresentação. Uma maneira eficiente de fazer isso é chamada de **multidimensionalidade**. A maior vantagem da multidimensionalidade é que ela permite que os dados sejam organizados como cada gerente gosta de vê-los, e não como os analistas de sistemas gostam. Apresentações diferentes dos mesmos dados podem ser providenciadas de modo rápido e fácil.

Apresentação multidimensional

As tabelas das planilhas têm duas dimensões. Informações com três ou mais dimensões podem ser apresentadas com o uso de um conjunto de tabelas de duas dimensões ou de uma tabela razoavelmente complexa. No suporte à decisão, é feito um esforço a fim de simplificar a apresentação de informações e permitir ao usuário alterar, de modo rápido e fácil, a estrutura das tabelas para torná-las mais expressivas (p.ex., ao inverter colunas e linhas, agrupar a sintetização de várias linhas e colunas, separar um conjunto de linhas ou colunas).

Três fatores são levados em consideração na multidimensionalidade: *dimensões*, *medidas* e *tempo*. Seguem alguns exemplos:

- *Dimensões*. Alguns exemplos de dimensões são: produtos, equipe de vendas, segmentos de mercado, unidades de negócio, localizações geográficas, canais de distribuição, países e setores.
- *Medidas*. Alguns exemplos de medidas são: dinheiro, volume de vendas, número de funcionários, estoque, e lucro previsto vs. real.
- *Tempo*. Alguns exemplos de tempo são: diário, semanal, mensal, trimestral e anual.

Um **banco de dados multidimensional** é um banco de dados onde os dados são organizados especificamente para permitir uma análise multidimensional rápida e fácil. Os dados são trazidos do *data warehouse*. Criar e manter esse tipo de banco de dados agrega custos que devem ser comparados aos benefícios agregados.

Aqui está um exemplo de como a multidimensionalidade funciona: um gerente deseja saber as vendas de um produto (por unidade ou dólar) em uma área geográfica específica, de uma equipe de vendas específica, durante um mês específico. A resposta para essa pergunta pode ser obtida muito mais rápido pelo próprio usuário, independentemente da estrutura do banco de dados, se os dados estiverem organizados em bancos de dados multidimensionais ou se a consulta ou produtos de *software* relacionados estiverem projetados para multidimensionalidade. Em ambos os casos, os usuários podem navegar através de muitas dimensões e níveis de dados por meio de tabelas ou gráficos e fazer interpretações rápidas, como, por exemplo, descobrir variações significativas ou tendências importantes.

Cubos de dados multidimensionais e suas análises

Um cubo de dados é usado para representar dados em conjunto com alguma medida de interesse. Embora seja chamado de “cubo”, pode ser bidimensional, tridimensional ou com uma dimensão superior. Cada dimensão representa algum atributo no banco de dados, e as células no cubo de dados representam medidas de interesse.

Exemplo

Um banco de dados contém informações sobre operações relacionadas às vendas de produtos (*p*) de uma empresa a um cliente (*c*) em diferentes localizações da loja (*s*). O cubo de dados formado a partir desse banco de dados é uma representação tridimensional, com cada célula (isto é, *p*, *c*, *s*) do cubo representando uma combinação de valores para *produto*, *cliente* e *localização da loja*. Um cubo de dados modelo para essa combinação é mostrado na Figura 3.3. O conteúdo de cada célula é a contagem do número de vezes que a combinação de valores ocorreu junta no banco de dados. As células que aparecem em branco equivalem a zero. O cubo pode ser usado, por exemplo, para acessar informações dentro do banco de dados para saber qual loja deve receber um determinado produto para venda a fim de obter o máximo de lucro.

Os cubos de dados oferecem a oportunidade de obter informações para suporte à decisão de maneira eficiente. A *análise de cubo* permite às pessoas realizar consultas através da busca de uma série de visualizações de relatórios, usando os recursos do *software* OLAP, como *page-by*, *pivot*, classificação, filtro e *drill-up/drill-down*, que permitem aos usuários fazer “visões parciais” de um cubo de dados ou analisar um cubo usando um simples clique do *mouse*. O termo **cubo** refere-se a um conjunto de dados altamente correlacionados que são organizados para permitir que os usuários combinem qualquer atributo em um cubo (p.ex., lojas, produtos, clientes, fornecedores) com qualquer métrica no cubo (p.ex., vendas, lucro, unidades, época) a fim de criar diversas visões bidimensionais, ou cubos, que podem ser exibidas em uma tela de computador (veja Figura 3.4).

Para obter informações sobre como a análise de cubo é feita, acesse o *site* ou confira os produtos de fornecedores, como Intelligent Cubes da MicroStrategy (MicroStrategy, 2005).

A multidimensionalidade está disponível em diferentes níveis de sofisticação. Portanto, há diversos tipos de *softwares* comerciais (com diferentes níveis de preço) a partir dos quais se podem criar sistemas multidimensionais.

Ferramentas multidimensionais e seus fornecedores

Muitas vezes, as ferramentas com recursos multidimensionais trabalham em conjunto com sistemas de consulta em bancos de dados e outras ferramentas OLAP. Por exemplo, o Crystal Reports da Seagate Software (pertencente à Seagate Technology LLC, seagate.com) cria relatórios que extraem e analisam dados provenientes do banco de dados relacional. O Crystal Analysis Professional cria relatórios que extraem e analisam dados multidimensionais provenientes de sistemas OLAP (p.ex., Hyperion Essbase), assim como de bancos de dados usuais com tecnologia OLAP integrada (p.ex., Microsoft SQL Server 2000, IBM DB2). Veja Hoffer et al. (2005). Outras ferramentas incluem o PowerPlay (cognos.com), o InterNetivity Databeacon (databeacon.com) e o Business Objects (businessobjects.com).

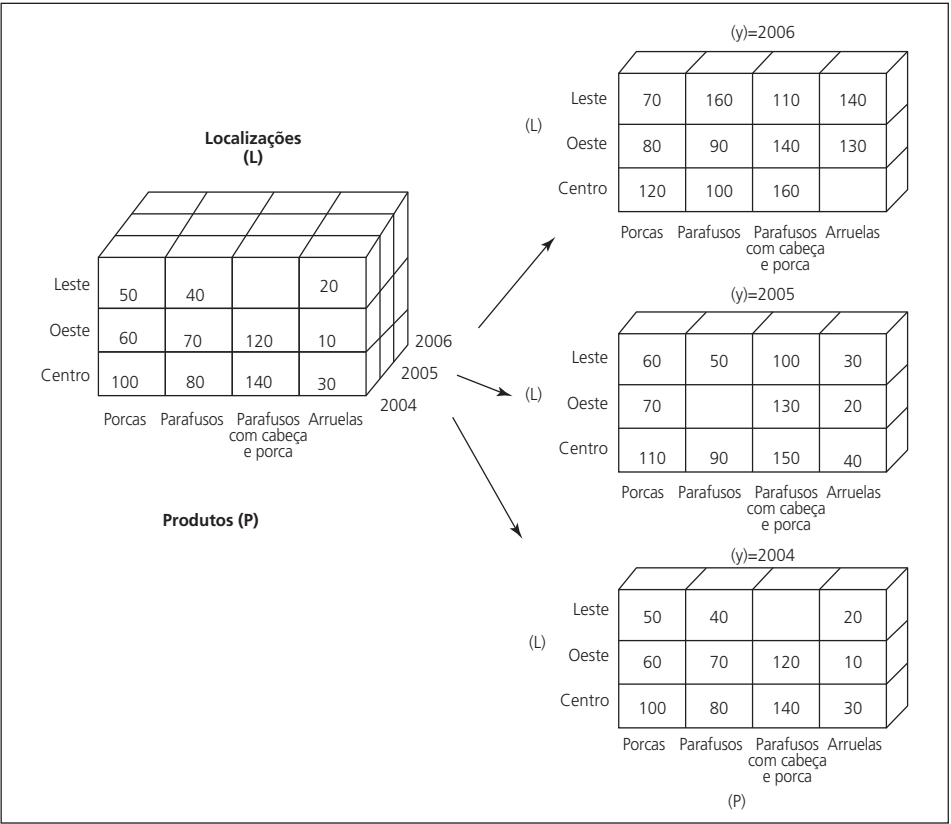


Figura 3.3 Visões e análise de cubo.

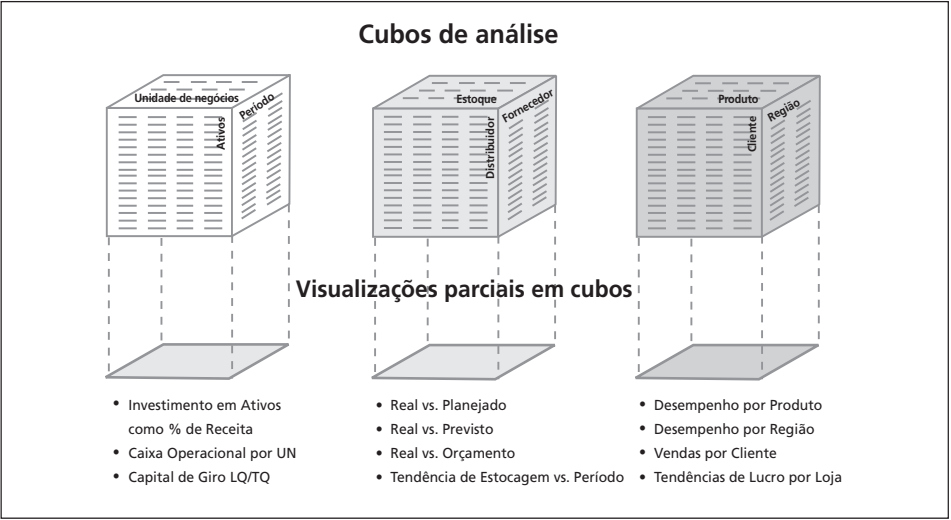


Figura 3.4 Visões parciais em cubos.

Fonte: MicroStrategy. "The Five Styles of Business Intelligence". 2003, Figura 21, p.35".

O Cube Views da IBM automatiza a criação de metadados OLAP no nível do banco de dados, de modo que esses metadados sejam compartilhados entre as aplicações que acessam o banco de dados. O Cube Views reúne dados em gráficos multidimensionais, permitindo aos usuários acessar os dados de diferentes perspectivas. Ele retorna respostas a consultas com *Web services* baseados em XML. Muitos fornecedores de BI, incluindo Hyperion, Cognos Inc., MicroStrategy Inc., Informatica Corp., InterNetivity e BusinessObject, dão suporte ao Cube Views. Consulte Callaghan (2003b) para mais detalhes.

Para exemplos de *softwares* de BI que executam a multidimensionalidade sem dificuldade, consulte Callaghan (2003b) e a edição “Annual Product Review” da *DM Review* todo mês de julho (dmreview.com).

Limitações da dimensionalidade

Segundo o relatório de uma pesquisa do Grupo Gartner (citado por Gray e Watson, 1998), a multidimensionalidade possui algumas limitações:

- O banco de dados multidimensional pode ocupar significativamente mais espaço de memória de um computador do que um banco de dados relacional compactado.
- Produtos multidimensionais custam significativamente mais do que produtos relacionais padrão.
- O carregamento do banco de dados consome tempo e recursos significativos do sistema, dependendo do volume de dados e número de dimensões.
- As interfaces e manutenção são mais complexas em bancos de dados multidimensionais do que em bancos de dados relacionais.

Essas limitações tornam-se problemas menores com o passar do tempo devido ao aperfeiçoamento da tecnologia.

Questões de revisão da Seção 3.5

1. Defina *multidimensionalidade* e descreva seus benefícios.
2. Defina *banco de dados multidimensional* e descreva as limitações desse tipo de banco de dados.
3. Defina *cubo* e *cubo de dados*.
4. O que significa visões parciais em cubos?
5. Liste as limitações da multidimensionalidade.

3.6

ANÁLISE DE NEGÓCIOS AVANÇADA

Enquanto o OLAP concentra-se em relatórios e consultas, outras análises têm sido desenvolvidas para examinar dados e informações de maneiras mais sofisticadas. Na realidade, empresas em todo o mundo usam centenas de modelos matemáticos, financeiros, estatísticos e outros para resolver problemas, explorar oportunidades, melhorar a produtividade e obter vantagens estratégicas.

Um exemplo da capacidade da BA pode ser visto no caso da KeySpan, uma distribuidora de gás natural no nordeste dos EUA. Os clientes comerciais da empresa, os comerciantes que vendem gás aos usuários finais, podem acessar a *extranet* da KeySpan, visualizar as projeções do uso de gás e tomar decisões informadas sobre o volume de gás que deve ser transportado pelas tubulações em qualquer dia específico. A KeySpan utiliza a tecnologia da MicroStrategy para analisar fatores como dados históricos e condições meteorológicas a fim de projetar o volume de gás que deve ser distribuído. Quatro vezes ao dia, o MicroStrategy Narrowcast Server compara o fluxo real de gás com a projeção feita no início do dia e, então, envia um relatório de exceção por *e-mail* ou *pager* ao comerciante e ao departamento de operações de gás que controla as válvulas. Os alertas avisam

os comerciantes quando eles projetaram o uso de gás a mais ou a menos para que possam fazer os ajustes necessários.

Nas seções seguintes, apresentaremos apenas alguns exemplos significativos de BA avançada. (Para mais exemplos, consulte histórias de sucesso do cliente em sas.com e spss.com.) Usuários podem realizar análises estatísticas e matemáticas sofisticadas, como teste de hipóteses, regressão e correlação múltiplas, previsões de dinâmica e modelos para pontuação de cliente. Essa investigação não pode ser feita com OLAP básico e requer ferramentas especiais, incluindo *data mining* e análise preditiva.

Data mining e análise preditiva

Um passo importante na tomada de decisão gerencial é prever ou estimar os resultados de procedimentos alternativos diferentes. Para isso, podemos usar um ou mais métodos. O método mais simples envolve somente duas variáveis, como, por exemplo, *vendas sobre tempo* ou *preço sobre tempo*. De fato, muitas fórmulas estatísticas estão disponíveis para realização dessas análises (p.ex., regressão e correlação, análise de tendências). As fórmulas para esses métodos aparecem como funções no Excel ou MicroStrategy.

Em muitos casos, porém, uma previsão é mais complexa, envolvendo mais de duas variáveis. Nesses casos, podemos usar métodos estatísticos mais complexos, como análise de regressão múltipla ou métodos de previsão especiais. Dois desses métodos são descritos a seguir: *data mining* e análise preditiva.

Data mining

As ferramentas de *data mining* extraem automaticamente, a partir de bancos de dados, informações preditivas escondidas. Elas também procuram padrões em bancos de dados de operações amplas. Essas ferramentas normalmente são baseadas em fórmulas estatísticas complexas. A maneira mais fácil de diferenciar *data mining* de outras formas de OLAP é que o OLAP consegue responder perguntas que se tem certeza de perguntar, enquanto que o *data mining* responde perguntas que não necessariamente se saiba que se deve perguntar. Os detalhes sobre *data mining* são fornecidos no Capítulo 4 e Capítulo online 6.

Análise preditiva

As ferramentas de **análise preditiva** ajudam a determinar o resultado provável futuro para um evento ou a probabilidade de uma situação ocorrer. Elas também identificam relações e padrões. Um exemplo é fornecido no Caso de aplicação 3.3.

CASO DE APLICAÇÃO 3.3

A análise preditiva pode ajudá-lo a evitar engarrafamentos

Agora, a análise preditiva pode ser usada para prever os níveis de engarrafamento com horas ou até dias de antecedência, e com quase 90% de precisão. A Inrix (inrix.com) é uma empresa iniciante (*startup*) que fornece esse tipo de previsão ao custo de \$20 a \$120 por ano. A análise preditiva é feita com um grande número de dados obtidos de fontes do governo, incluindo:

- Informações sobre fluxo de tráfego e incidentes em tempo real recolhidas por dispositivos eletrônicos instalados nas rodovias (leitores de placa, câmeras, radares e sensores magnéticos embutidos no pavimento)
- Dados de localização e velocidade recolhidos por unidades de sistema de posicionamento glo-

bal (GPS) em veículos pertencentes às empresas de transporte e entrega participantes

- Dois anos de dados históricos do fluxo de tráfego
- Previsões e condições meteorológicas
- Outros eventos (p.ex., cronogramas de construção de estrada, calendários escolares, esportes, concertos, outros eventos especiais agendados)

Os algoritmos preditivos de propriedade exclusiva da Inrix combinam esses dados para criar um panorama do fluxo de tráfego atual, do engarrafamento esperado e das condições da estrada ao longo das próximas horas e dias. Evidentemente, cada cidade requer seu próprio modelo e banco de dados (0,2 a 2 *terabytes* por cidade). Em 2006, a Inrix ofereceu essa previsão em 30 cidades. O serviço é combinado com mapas digitais (acesse teleatlas.com, o provedor de informações para empresas de GIS como a MapQuest). A Inrix também se associa a operadoras de telefone celular, transmissores de satélite tradicionais e serviços de navegação para carros. Na região de Seattle, onde a Inrix está localizada, a empresa fornece informações sobre o tráfego via *smartphones* e painéis eletrônicos em trechos da estrada, usando códigos coloridos para sinalização. Os *smartphones*

também exibem o tempo estimado para as estradas ficarem livres ou congestionadas.

O sistema da Inrix sugere decisões automáticas, como por exemplo:

- Melhor trajeto para uma *van* de entrega
- Tempo ideal para ir ou sair do trabalho (para aqueles que possuem um horário flexível)
- Como refazer um trajeto para evitar um acidente

Os seguintes exemplos são algumas das tecnologias em uso para sensoriamento e controle do tráfego:

- Laços detectores magnéticos na superfície da estrada (em 32% das cidades dos EUA)
- Câmeras de circuito fechado de TV monitorando as condições do tráfego (em 29% das cidades dos EUA)
- Informações sobre condições do tráfego fornecidas por rádio e na Internet em tempo real (em 19% das cidades dos EUA)
- Rampas de acesso à auto-estrada controladas por semáforos (em 9% das cidades dos EUA)

Fontes: compilado de E. Jonietz, “Traffic Avoidance”, *Technology Review*, December 2005/January 2006; and J. Barke, “Traffic Taming”, *Technology Review*, October 3, 2005.

Segundo Fogarty (2004), a análise preditiva usa algoritmos sofisticados criados para examinar um *data warehouse* e identificar padrões de comportamento que sugerem, por exemplo, a quais ofertas seus clientes poderiam ser suscetíveis no futuro ou quais clientes há o perigo de você perder. Ao examinar um *data warehouse* de um banco, uma análise preditiva poderia confirmar que clientes que cancelam um pagamento em débito automático ou depósito automático frequentemente mudam para outro banco dentro de um determinado período de tempo.

Análise preditiva pode examinar padrões de atividade em contas a fim de identificar localizações geográficas ou mudanças e, então, conseguir correlacionar esses fatores para encontrar padrões ligados às contas que foram encerradas. Essa abordagem identificaria não apenas que um pagamento automático cancelado era um mau sinal, mas também que a localização não era um fator relevante. O exemplo de análise preditiva a seguir é fornecido por McKinley (2004).

Exemplo: reconhecer o que os clientes desejam antes mesmo de eles entrarem no restaurante

A HyperActive Technologies (hyperactivetechnologies.com) desenvolveu um sistema no qual câmeras instaladas no telhado de um restaurante *fast-food* controlam os veículos que param no estacionamento (ou usam o *drive-thru*). Outras câmeras controlam o progresso dos clientes na fila de pedidos. Usando a análise preditiva, o sistema prevê o que os clientes poderiam pedir. Além disso, um banco de dados inclui dados históricos de pedidos feitos no carro, como, por exemplo, 20% dos carros que entram no estacionamento geralmente pedirão, pelo menos, um *cheeseburger* na hora do almoço. Com base nas informações das câmeras e no banco de dados, o sistema prevê o que os clientes pedirão 1,5 a 5 minutos antes de eles de fato pedirem. Isso dá aos cozinheiros a chance de preparar a comida, minimizando o tempo de espera do cliente. Além disso, não há a possibilidade de a comida esfriar (reaquecer leva tempo, custa dinheiro e torna a comida menos saborosa).

A análise preditiva é usada para examinar automaticamente grandes quantidades de dados com diferentes variáveis; incluindo *clustering*, árvores de decisão, análise de cesta de supermercado, modelagem de regressão, redes neurais, algoritmos genéticos, *text mining*, teste de hipóteses, análise de decisão e mais.

O elemento central da análise preditiva é o *indicador*, uma variável que pode ser medida por um indivíduo ou entidade a fim de prever um comportamento futuro. Uma empresa de cartão de crédito poderia, por exemplo, considerar a idade, renda, o histórico de crédito e outras estatísticas demográficas como indicadores na determinação do fator de risco de um solicitante. Para mais informações sobre análise preditiva, consulte Zaman (2005).

A segunda geração de produtos ERP e SCM é caracterizada pelo acréscimo de ferramentas de suporte à decisão/BI. Um exemplo é a análise SAP (veja o Arquivo *online* W3.1).

Ferramentas para análise avançada de fornecedores representativos

A lista a seguir mostra produtos específicos para análise de BI oferecidos pelos principais fornecedores do ramo.

- A *MicroStrategy* fornece mais de 400 funções estatísticas, matemáticas e financeiras para criação de relatórios e análise de seus resultados. Isso inclui ferramentas como *drill-anywhere*, divisão de dados, agrupamento de dados e otimização.
- O *System 9 BI* da *Hyperion* inclui o módulo Essbase Analytics para realizar rapidamente análises sofisticadas que permitem aos gerentes interpretar dados complexos. Módulos especiais são reservados para parte financeira, análise empresarial, padrões de comportamento do cliente, redução de custos, conformidade e desempenho corporativo.
- A análise do *Cognos 8 Business Intelligence* inclui análise de séries temporais e tendências personalizáveis, análise profunda da concorrência, detalhamento, previsão e otimização. A *Cognos* também oferece gerenciamento de eventos de negócios.
- A *Microsoft* oferece análise avançada no seu *Microsoft Dynamics GP 9.0*.
- A *Fair Isaac* oferece ferramentas inteligentes para realização de análise de risco, detecção de fraude, análise de lucratividade e consulta inteligente.
- A *ILOG* oferece um pacote de otimização sofisticado. Com otimização, é possível maximizar a utilização de recursos, descobrir opções para solucionar problemas e realizar análises de custo-benefício.
- A *SAS* é líder em análise preditiva e *data mining* (p.ex., *SAS Enterprise Miner*). Ela fornece ferramentas financeiras, estatísticas e de previsão para a solução de problemas. Uma área de destaque é a de avaliação e gerenciamento de risco, usando modelos de pontuação de crédito (*credit scoring*).
- A *SPSS* é outra líder em análise preditiva e *data mining*. Conhecida há anos por seus pacotes estatísticos sofisticados, a *SPSS* está mais voltada para a análise preditiva atualmente.
- A *Oracle* fornece diversas ferramentas para análise preditiva, *data mining* e outras ferramentas para análise avançada.
- Também estão nesse grupo *Insightful Corp.*, *StatSoft Inc.*, *Knowledge eXtraction ENgines*, *Unica*, *Angoss Software* e *IBM*.

Questões de revisão da Seção 3.6

1. Por que as ferramentas de análise avançada são necessárias para as empresas?
2. Liste algumas ferramentas de BA avançada.
3. Defina *data mining*.
4. Defina *análise preditiva* e descreva seus recursos.
5. Liste algumas ferramentas de análise preditiva.

VISUALIZAÇÃO DE DADOS

As tecnologias visuais podem condensar 1.000 números em uma única imagem e tornar as aplicações de suporte à decisão mais atraentes e compreensíveis aos usuários. **Visualização de dados** refere-se às tecnologias que dão suporte à visualização e, algumas vezes, à interpretação de dados e informações em vários pontos ao longo da cadeia de processamento de dados (veja Fayyad et al., 2002). Ela inclui imagens digitais, sistemas geográficos, interfaces gráficas de usuário, gráficos, realidade virtual, representações dimensionais, vídeos e animações. As ferramentas visuais podem ajudar a identificar relações, como por exemplo, tendências.

A capacidade de identificar de maneira rápida tendências importantes em dados corporativos e de mercado pode fornecer vantagem competitiva. Quando reconhecemos tendências visualmente, podemos verificar a importância destas usando modelos preditivos que fornecem vantagens comerciais significativas em aplicações que mudam conteúdo, transações ou processos. Uma ação segura, baseada em métodos superiores de análise visual de dados, ajuda as empresas a melhorar a renda e evitar erros caros (veja Hallet, 2001). A Consonus (**consonus.com**), por exemplo, projeta, cria e opera centros de dados, redes de TI e sistemas de entrega de aplicações capacitadas para Web. A Consonus usa o pacote de gerenciamento OpenView da Hewlett-Packard para ajudar a gerenciar esses centros de dados. O OpenView também ajuda a gerenciar sistemas Web e fornece aos usuários percepção de como os clientes vêem o desempenho e a disponibilidade dos seus *sites*. Consulte McCarthy e McCarthy (2002) para mais detalhes.

A visualização de dados também habilita BA, ao usar ferramentas baseadas na Web. Em vez de ter de esperar um relatório ou comparar colunas inúteis de números, um gerente pode usar uma interface de navegador em tempo real para analisar dados vitais de desempenho organizacional. Ao usar tecnologias visuais de análise, gerentes, engenheiros e outros profissionais podem reconhecer problemas que passaram despercebidos, durante anos, pelos métodos de análise padrão.

A visualização de dados é mais fácil de implementar quando os dados necessários estão em um *data warehouse* ou, melhor ainda, em um banco de dados multidimensional especial ou servidor. Um exemplo é a Harrah's Entertainment (veja o caso da Harrah's em **teradatauniversitynetwork.com**), que instalou a tecnologia de visualização Compudigm International's (**compudigm.com**) nas suas matrizes em Las Vegas e em outros cassinos. Um ambiente de visualização apresenta dados de uma maneira que possibilita aos tomadores de decisão ver profundidade e qualidade em tempo real e influenciar o desempenho. Agora, os tomadores de decisão da Harrah's podem ver o fluxo de tráfego em toda a área do cassino em tempo real. Eles podem identificar quais máquinas caça-níqueis são as mais procuradas pelos clientes e quais são as mais rentáveis, por minuto, e podem instalar mais das melhores máquinas, quando for necessário. O *software* de visualização de dados permite aos gerentes determinar a disposição do cassino de forma dinâmica e também examinar visualmente seus programas de fidelidade do CRM. Inicialmente, a Compudigm desenvolveu esse produto para o setor de jogos de azar, mas desde então tem ampliado sua tecnologia para serviços financeiros e telecomunicações.

Visualização por planilhas

As planilhas são as principais ferramentas do usuário final para programação de aplicações de suporte à decisão. O Microsoft Excel oferece dúzias de ferramentas matemáticas, estatísticas, de geração de relatório, de consulta (p.ex., análise de hipóteses, identificação de objetivos) e outras ferramentas de BI.

O Excel tem sido amplamente adotado como uma ferramenta eficiente e fácil de usar para manipulação de dados em formato livre. Ele tem evoluído para além de uma simples ferramenta

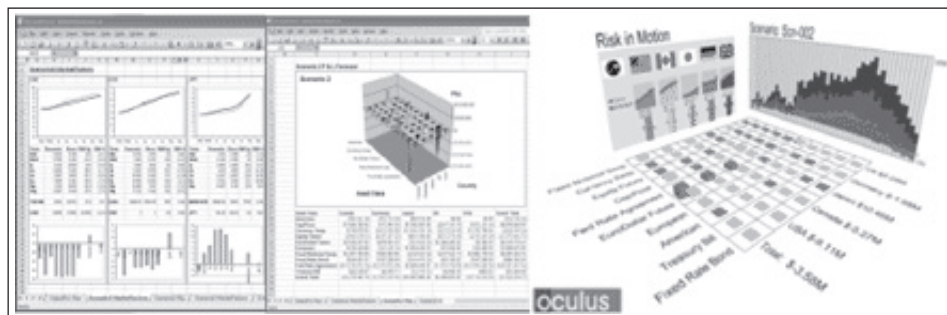
Exemplo: Gerenciamento de risco

Exemplo: Compreensão dos modelos MSS

Muitos outros exemplos de visualização de planilhas e suas vantagens são fornecidos por Brath e Peters (2006).

Desde o fim dos anos 90, a visualização de dados evoluiu tanto na computação convencional, onde é integrada às ferramentas e aplicações de suporte à decisão, como na visualização inteligente, que inclui a interpretação de dados (informação). Para tendências e áreas interessantes, veja o Arquivo *online* W3.3.

Os principais fornecedores de OLAP oferecem ferramentas de visualização tridimensional junto com suas ferramentas de suporte à decisão. O Forest Tree 6.0, por exemplo, é uma ferramenta de desenvolvimento que tem uma versão para visualização tridimensional que permite aos



Fonte: R. Brath and M. Peters, "Visualizing Spreadsheets: Add Exponential Value to Excel for Monitoring, Analysis and Modeling", *DM Direct Newsletter*, January 6, 2006.

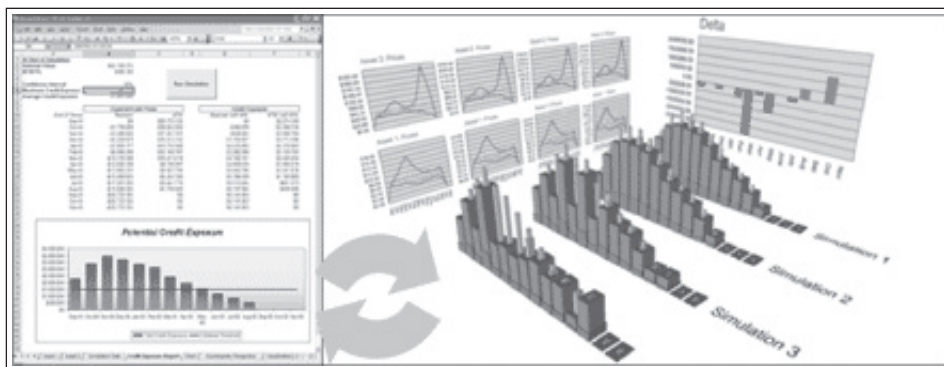


Figura 3.6 Planilha visual de modelagem de crédito.

Fonte: R. Brath and M. Peters, “Visualizing Spreadsheets: Add Exponential Value to Excel for Monitoring, Analysis and Modeling”, *DM Direct Newsletter*, January 6, 2006.

usuários ver e gerenciar facilmente múltiplas dimensões de dados em uma única vista. Novas ferramentas visuais são desenvolvidas continuamente para analisar o desempenho de *websites*. A ADVIZOR Solutions (advizorsolutions.com) é uma dessas ferramentas.

Dashboards e indicadores

É comprovado que a visualização é extremamente importante para executivos atarefados. O sistema de informações gerenciais (EIS) dos anos 90 era cheio de gráficos e tabelas. Ele evoluiu para produtos de gerenciamento de *cockpit* (veja sap.com) e, posteriormente, para *dashboards* e indicadores (veja o Capítulo 5).

Análise visual

Atualmente, a análise de dados empresariais pode ser feita por usuários não-técnicos que colhem informações valiosas provenientes de dados comerciais. A VizQL (da Tableau Software) é uma linguagem visual de consulta a banco de dados que ativa o Hyperion Visual Explorer. Diversas outras empresas fornecem ferramentas para análise visual (p.ex., Analytica [lumina.com], Endeca [endeca.com]). A análise visual pode ser feita de maneira interativa (p.ex., veja navteq.com). A seguir, é descrita uma área da análise visual.

Visualização de dados financeiros

Uma das aplicações freqüentes da visualização de dados de BI é na área financeira. Para evitar que sistemas identifiquem automaticamente padrões inexpressivos nos dados, os diretores financeiros (CFOs) querem ter certeza de que a capacidade de processamento de um computador sempre será ajustada pelo discernimento de um ser humano. Uma maneira de fazer isso é através da visualização de dados. Diretor do Laboratório de Engenharia Financeira da Sloan School of Management do MIT, Andrew W. Lo desenvolveu um programa no qual um CFO pode usar o *mouse* para “voar” sobre um panorama tridimensional que representa o *risco*, *lucro* e a *liquidez* dos ativos de uma empresa. Com prática, o CFO pode começar a se concentrar no ponto selecionado no panorama tridimensional: aquele onde o equilíbrio entre risco, lucro e liquidez é o mais vantajoso.

Os CFOs têm usado a visualização tridimensional de modo crescente no ciberespaço, mas muitos ainda gastam a maior parte do seu tempo em assuntos de rotina, como geração de relatórios para a Comissão de Valores Mobiliários, usando as ferramentas tradicionais. Entretanto, essa situação está mudando, conforme descrito no Caso de aplicação 3.4.

CASO DE APLICAÇÃO 3.4**A visualização de dados financeiros na Merrill Lynch**

Os mapas baseados na Web da SmartMoney.com fornecem aos analistas da Merrill Lynch uma representação visual das informações financeiras fácil de examinar e imagens tridimensionais desses mercados em constante mudança. Para prestadores de serviços financeiros como a Merrill Lynch, a visualização está se tornando onipresente em muitas aplicações; além disso, os desenvolvedores estão ampliando seu uso para atender, cada vez mais, públicos variados. Do setor de serviços financeiros às empresas de engenharia de qualidade altamente técnicas, que criam ferramentas de teste, e aos prestadores de serviço, que usam ferramentas de monitoramento de rede eficazes, a visualização está se tornando uma ferramenta de valor agregado sob demanda. Os usos futuros dessa tecnologia são planejados e evoluem à medida que os pesquisadores descobrem aplicações para os negócios de amanhã.

O setor de serviços financeiros é um sólido adiante dessas novas ferramentas, com um mercado sofisticado crescente e um alto nível de concorrência.

Para responder a algumas dessas complexidades de dados, a Merrill Lynch usa mapas para seus produtos de dados internos patenteados a fim de acessar atualizações em tempo real nas ações e nos fundos mútuos.

Os corretores da Merrill Lynch podem controlar as posses individualmente ou no todo, recebendo panoramas de investimento de seus próprios portfólios e dos portfólios dos clientes. A tecnologia de mapeamento pode apresentar dados (p.ex., através de cores) com volume, valor e hierarquia, e permite aos usuários clicar em e acessar grupos de dados específicos a partir do mapa. Os usuários podem criar árvores tridimensionais dinâmicas e interativas a partir dos relatórios financeiros da empresa, compartilhá-las com os usuários através da Internet e integrá-las às aplicações existentes.

Fontes: compilado de J. McCarthy, “Envisioning Enterprise Data”, *InfoWorld*, November 18, 2002; and H. Goldberg, “Unleashing the Power of Data”, *DBZ Magazine*, 2º trimestre, 2004.

Fornecedores e produtos

Os pacotes de *software* de visualização oferecem aos usuários recursos para exploração autoorientada e análise visual de grandes volumes de dados. Veja o pacote de visualização da ILOG (ilog.com). Alguns exemplos de sistemas OLAP que oferecem excelente visualização incluem eBizinsights V(isual Insights) e nVizn (SPSS Inc.). O Cognos Visualizer usa exibições de semáforo em tabelas e gráficos. Os resultados numéricos são exibidos em vermelho, amarelo e verde para indicar seus status. O nVizn da SPSS é um *kit* de ferramentas para desenvolvedor baseado em Java para criação de aplicações de visualização. As tecnologias de visualização podem ser integradas para criar diferentes apresentações de informação, especialmente aos métodos de realidade virtual (RV) (veja Arquivo *online* W3.4). O Executive Viewer (da Temtec) também usa a idéia de semáforo na exibição de dados e integra os dados a uma estrutura multidimensional (veja o Exercício 1 em “Tarefas em grupo e interpretação de papéis”, no fim deste capítulo).

Questões de revisão da Seção 3.7

1. Defina *visualização de dados* e liste suas principais vantagens.
2. Descreva visualização de planilha.
3. Descreva visualização de dados financeiros.

3.8**SISTEMAS DE INFORMAÇÃO GEOGRÁFICA (GIS)**

Um **sistema de informações geográficas (GIS)** é um sistema baseado em computador para captura, armazenamento, modelagem, recuperação, verificação, integração, manipulação, análise e exibição de dados citados geograficamente através do uso de mapas digitais. Para mais informa-

ções sobre mapas digitais, acesse gpsy.com/maps. A característica mais distintiva do GIS é que cada registro ou objeto digital tem uma localização geográfica identificada. Ao integrar mapas aos bancos de dados orientados espacialmente (ou seja, localização geográfica) (chamados de *geocodificação*) e a outros bancos de dados, os usuários podem gerar informações para planejamento, resolução de problemas e tomada de decisão com isso aumentar sua produtividade e a qualidade das suas decisões. Áreas tão heterogêneas como varejo, bancos, transportes, agricultura, gestão de recursos naturais, administração pública, controle do espaço aéreo, militar, prontidão de emergência e planejamento urbano têm usado o GIS com êxito desde o início dos anos 70 (veja Ursery, 2004). Para fontes de dados geográficos, consulte Saarenvirta (2004).

Aplicações de GIS

As ferramentas de GIS tornam-se cada vez mais sofisticadas e acessíveis. Elas ajudam as empresas e os governos a saber onde precisamente estão seus caminhões, funcionários e recursos; para onde eles precisam ir para atender um cliente; e a melhor maneira para deslocar-se. As áreas de *marketing* direcionado estão crescendo rapidamente, e as empresas podem segmentar facilmente uma população usando o GIS. A Cooperativa de Crédito do Texas (cuoftexas.com), por exemplo, usa o GIS para ajudar a decidir onde colocar *outdoors* e caixas automáticos e ajudar a identificar as áreas mais receptivas para mala direta. A taxa de resposta característica para essa cooperativa de crédito varia entre 5 e 10%, muito melhor do que a tradicional média de 1 a 2%. Os clientes também apreciam receber menos correspondências da cooperativa de crédito, pois recebem apenas correspondências direcionadas. Para mais detalhes, consulte Franklin (2002) e Ursery (2004).

Exemplo

O estado da Louisiana está usando o WebFOCUS da Information Builders (informationbuilders.com/products/webfocus/index.html) para identificar pessoas que comercializam ilegalmente vales-alimentação. Também usa o GIS em seus departamentos de polícia para descobrir padrões geográficos em crimes, assim como para posicionar policiais. Também usa o GIS no gerenciamento de ativos físicos. Para mais detalhes, consulte Schwartz (2005).

Para muitas empresas, a organização inteligente de dados dentro de um GIS pode oferecer uma estrutura de suporte a processos de tomada de decisão e de definição de estratégias alternativas. Alguns exemplos de aplicações de GIS bem-sucedidas estão resumidas na Tabela 3.5. Empresas de destaque incorporam o GIS aos seus sistemas de BI. O GIS incorpora, *em condições ideais*, dados de censo (veja census.gov) como fonte de dados demográficos para tomada de decisão eficaz. Para muitas organizações, o GIS e as análises espaciais relacionadas são uma prioridade principal. A Sears, por exemplo, investiu milhares de dólares em tecnologia de GIS na área logística, levando a uma economia de US\$ 52 milhões por ano (veja Gonzales, 2003).

Muitos bancos usam o GIS para dar suporte a atividades como:

- Determinação da localização de agências e caixas automáticos
- Análise da faixa demográfica (p.ex., residência, idade, nível de renda) do cliente para cada um dos produtos do banco
- Análise dos padrões de volume e tráfego das atividades comerciais
- Análise da área geográfica atendida pela agência
- Determinação de mercado potencial para atividades bancárias
- Avaliação dos pontos fortes e fracos em relação aos da concorrência
- Avaliação do desempenho da agência

Os bancos também usam o GIS como planilhas geográficas que permitem aos gerentes modelar as atividades comerciais e realizar análises de hipóteses (p.ex., e se nós fecharmos uma agência

Tabela 3.5 Aplicações de GIS

<i>Empresa</i>	<i>Aplicação do GIS</i>
Pepsi Cola Inc., Super Value, Acordia Inc.	Usam o GIS na seleção de locais para novos restaurantes Taco Bell e Pizza Hut. Combinam dados demográficos e padrões de tráfego.
CIGNA (seguro de saúde)	Usa o GIS para responder perguntas como: quantos médicos afiliados ao CIGNA estão disponíveis em um raio de 13 km de uma empresa?
Western Auto (uma subsidiária da Sears)	Integra dados ao GIS para criar um perfil demográfico detalhado da região de uma loja a fim de determinar a melhor combinação de produtos a serem oferecidos na loja.
Sears, Roebuck & Co./Kmart	Usa o GIS para dar suporte ao planejamento das rotas rodoviárias.
Empresas de planos de saúde	Controlam as incidências de câncer e de outras doenças a fim de determinar estratégias de expansão e alocação de equipamentos caros nas suas instalações.
Wood Personnel Services (agência de emprego)	Mapeia regiões onde moram trabalhadores temporários para localizar cidades para <i>marketing</i> e recrutamento.
Wilkening & Co. (serviços de consultoria)	Planeja áreas e rotas de vendas favoráveis para seus clientes, reduzindo as despesas de viagem em 15%.
CellularOne Corporation	Mapeia toda sua rede de celulares para identificar grupos de desconexão de ligações e enviar técnicos de maneira apropriada.
Sun Microsystems	Gerencia propriedades alugadas em dezenas de lugares no mundo todo.
Consolidated Rail Corporation	Monitora a condição de mais de 32.000 km de estrada de ferro e milhares de lotes de terra adjacente.
Secretaria de Defesa Civil Norte-Americana	Avalia danos causados por furacões, enchentes e outros desastres naturais através da relação dos vídeos dos danos com os mapas digitais das propriedades.
Toyota e outros fabricantes automotivos	Combinam GIS e GPS como ferramenta de navegação para orientar motoristas aos seus destinos através das melhores rotas.

ou unificarmos agências? E se um concorrente abrir uma agência?). Cada mapa consolida páginas de análise.

GIS e tomada de decisão

O GIS oferece uma grande quantidade de informações extremamente úteis que podem ser analisadas e utilizadas na tomada de decisão. O formato gráfico de um GIS facilita a visualização de dados pelos gerentes. Segundo Janet M. Hamilton, gerente de pesquisa de mercado da Dow Elanco, fabricante de defensivos agrícolas de US\$ 2 bilhões com sede em Indianápolis, “Posso colocar planilhas de 80 páginas com milhares de linhas em um único mapa. Levaria algumas semanas para compreender todas as informações da planilha, mas, em um mapa, a história pode ser contada em segundos” (Hamilton, 1996, p.21).

Há inúmeras aplicações de GIS para melhorar a tomada de decisão nos setores público e privado (veja Saarenvirta, 2004), incluindo o envio de veículos de emergência, supervisão do trânsito, seleção de local para instalações, gerenciamento de risco de estiação e controle de fauna e flora silvestres. Os GISs são extremamente populares entre os governos regionais, onde são usados não apenas para mapear, mas também em muitas outras aplicações de tomada de decisão (veja Ursery, 2004). Estados e municípios estão usando aplicações de GIS relacionadas à taxaço e mapeamento de propriedades, controle de enchentes, entre outros.

Saarenvirta (2004) pesquisou sobre o valor do GIS e sua relação com a BI. Ele descreveu análises espaciais em áreas como mapeamento de cliente, mapeamento temático, análise da área de comércio e localização de varejo.

GIS associado ao GPS

O Departamento de Defesa dos EUA investiu cerca de US\$21 bilhões em sistemas de satélite que alimentam **sistemas de posicionamento global (GPS)**. Os dispositivos GPS são *wireless* e usam satélites para permitir que os usuários detectem a posição na Terra dos itens nos quais os dispositivos estão anexados (p.ex., carros, pessoas), com precisão razoável (consulte trimble.com/gps). O GPS em conjunto com o GIS está trazendo grandes progressos nas aplicações de BI. São inúmeros os usos comerciais e governamentais, pois os dispositivos de detecção são relativamente baratos. Veja o Caso de aplicação 3.5 para obter exemplos de como essas tecnologias estão sendo usadas.

CASO DE APLICAÇÃO 3.5

GIS e GPS rastreiam onde você está e ajudam a fazer as coisas

A seguir estão exemplos de como o GIS em conjunto com o GPS ajuda as empresas a diferenciar seus produtos, entregar serviços e melhorar o desempenho:

- A UltraEx, empresa da costa oeste dos EUA que é especializada em entregas no mesmo dia de itens como estoques de sangue emergenciais e peças de computador, equipa todos seus veículos com receptores GPS e *modems* sem fio da @Road. Além de dar aos despachantes uma visão panorâmica de toda a frota, a @Road ajuda a UltraEx a manter seus clientes satisfeitos ao deixá-los rastrear a localização e velocidade de seus carregamentos em tempo real na Web. Esse serviço mostra aos clientes um mapa do último local onde o satélite detectou o veículo de entrega e a velocidade na qual está viajando. Os motoristas usam o Serviço de Dados Móveis da AT&T para se comunicar com os despachantes, e os motoristas que são donos dos veículos não conseguem falsificar as planilhas de quilometragem porque a @Road informa a quilometragem exata de cada veículo.
- A cidade de Nova Iorque foi pioneira no uso do CompStat, que usa o GIS para mapear atividades criminais e disposição policial por data, hora e local. Ao tornar os comandantes dos distritos policiais responsáveis pela própria estratégia de policiamento, o CompStat foi um dos principais fatores para redução de aproximadamente 70% na taxa de crimes violentos da cidade na última década.
- Alguns departamentos de polícia, associações de bairros e cidadãos interessados, em diversas cidades, estão usando o GIS para combater a criminalidade. Informações geográficas sobre os bairros são integradas aos relatórios de criminalidade para analisar os padrões de crimes. Ao identificar tendências e fornecer informações aos cidadãos, a polícia está mais bem preparada para criar atividades de vigilância, e os cidadãos podem modificar atitudes, levando a uma redução da criminalidade nessas áreas.
- A CSX Transportation Inc. equipou com GPS 3.700 locomotivas. A Union Pacific Railroad instalou dispositivos de monitoramento via satélite em milhares de seus vagões de carga para rastreamento. Ao combinar GIS com GPS, uma empresa de transporte de cargas pode identificar a posição de um vagão ou caminhão em até 100 metros, a qualquer hora. Pode identificar, por exemplo, locomotivas que saíram da sua rota e vagões que foram deixados para trás ou que foram enviados com a locomotiva errada. Um benefício adicional é a capacidade de minimizar acidentes.
- No National Environmentally Sound Production Agriculture Laboratory da Universidade de Geórgia, cientistas desenvolveram um trator que é controlado por um giroscópio e GPS. Mau tempo e problemas de visibilidade não são mais uma preocupação, pois o trator sabe aonde ir. Os cientistas desenvolveram dispositivos que permitem ao sistema detectar pequenos obstáculos usando um sistema de visão robótica e métodos de inteligência artificial para interpretar o que o robô vê.
- No comércio de localização (*l-commerce*), a propaganda é direcionada a uma pessoa cuja localização é conhecida (através de uma combinação de GIS e GPS). De forma semelhante, sistemas médicos de emergência identificam, em segundos, a localização de um acidente de carro, e o GIS vinculado ajuda a orientar as ambulâncias até o local do acidente. Para outras aplicações interessantes, consulte Turban et al. (2006).

Fontes: adaptado de A. Dragoon, "Putting IT on the Map", *CIO*, May 15, 2003, and L. Rosencrance, "Railroads Hot for Satellite Monitoring", *Computerworld*, April, 2000.

GIS e a Internet/Intranets

A grande maioria dos fornecedores de *software* GIS oferece acesso à Web, como navegadores incorporados ou um servidor Web/Internet/*intranet* que se conecta diretamente ao *software* deles. Dessa maneira, os usuários podem acessar mapas e dados dinâmicos através da Internet ou de uma *intranet* corporativa. Os serviços de GIS na Web estão multiplicando-se. Esses sistemas geográficos formam uma infra-estrutura global cheia de informações que adiciona uma nova dimensão aos campos de GIS ao integrar serviços de aplicação múltiplos e distintos. Os serviços de GIS na Web estão revolucionando como as empresas usam e interagem com a informação geoespacial. O GIS pode, por exemplo, ajudar o gerente de uma atividade varejista a determinar se deve abrir uma loja localizada em um grande cruzamento da cidade, 15 minutos de carro de uma rampa de saída para estrada, rodeada por bairros de classe média de famílias trabalhadoras (veja Gonzales, 2003). A Big Horn Computer Services (Buffalo, NY) usou um GIS adaptado para Web a fim de desenvolver uma aplicação personalizada para a rede nacional de televisão que queria que suas emissoras afiliadas pudessem acessar uma *intranet* com informações demográficas sobre seus espectadores. Agora, os funcionários em cada emissora podem visualizar tematicamente mapas sombreados para analisar seus mercados usando um navegador da Web.

Várias empresas estão implementando GIS na Internet para usos internos ou para uso dos seus clientes. A Visa Plus, que opera uma rede de caixas eletrônicos, desenvolveu uma aplicação de GIS que permite aos usuários da Internet acessar um mapa para localizar qualquer um dos 300.000 caixas eletrônicos da empresa no mundo todo. Muitos fornecedores (p.ex., **borders.com**) disponibilizam localizadores de lojas. Esse localizador fornecerá o endereço da loja mais próxima, com um mapa de como chegar lá. À medida que fornecedores implementam *softwares* de servidor Web GIS, mais aplicações serão desenvolvidas. Mapas, dados de GIS e informações sobre GIS estão disponíveis na Web através de vários fornecedores e órgãos públicos. Fontes do governo (através de Internet e CD-ROM) fornecem alguns dados, e os fornecedores também oferecem dados comerciais diversificados. Algumas dessas informações são grátis (veja os CD-ROMs da MapInfo e materiais passíveis de *download* em **esri.com** e **data.geocomm.com**). Relacionado a isso está a inclusão de dados espaciais em *data warehouses* para uso posterior com tecnologia da Web.

Alguns exemplos de *softwares* GIS importantes são ArcView e ArcInfo (ESRI), AGISMap (AGIS), GeoMedia (intergraph) e MapInfo Professional (MapInfo). O modelo de dados do ArcInfo (**esri.com/arcinfo8/**) oferece ferramentas para modelar sistemas espaciais complexos sem programação. Culpepper (2002) descreveu como o *software* CommunityViz (**communityviz.com**) integra a funcionalidade de simulação e modelagem para planejamento de cidades no *softwares* GIS ArcView da ESRI. O usuário pode configurar e executar diferentes cenários, com base em variáveis e limitações definidas pelo usuário, para determinar relações entre projetos municipais e indicadores sociais, ambientais ou econômicos. Os usuários podem testar conjuntos inteiros de normas.

Questões de revisão da Seção 3.8

1. Defina *GIS* e descreva seus benefícios.
2. Como os bancos usam o GIS?
3. Descreva as aplicações GIS/GPS.
4. Por que a combinação de GIS e GPS é tão útil?
5. Como o GIS é relacionado à Web/Internet?

Três tecnologias emergentes estão estreitamente relacionadas à BA: BI em tempo real, suporte automatizado à decisão (ADS) e inteligência competitiva.

BI em tempo real

Os usuários de negócios estão, cada vez mais, exigindo acesso a dados em tempo real, não-estruturados ou remotos, integrados ao conteúdo de um *data warehouse* (veja Devlin, 2003). Os ônibus em Houston, Texas, por exemplo, estão mais seguros e eficientes desde que foram equipados com dispositivos que coletam dados instantaneamente, dando aos controladores de tráfego a capacidade de acessar informações e alterar os intervalos dos semáforos (veja “Houston Buses Due for ‘Intellectual Overhaul’”, 2003). Em muitos casos, atualizações e acessos aos dados em tempo real são fundamentais para o sucesso ou até mesmo sobrevivência de uma empresa. Segundo Baer (2002), a cidade de Richmond, British Columbia, usa coleta e análise de dados em tempo real. Richmond está localizada em uma ilha costeira e tem uma elevação média de apenas 91 cm acima do nível do mar. É importante para as autoridades da cidade saber imediatamente se a rede de bombas para controle da maré estão funcionando, o quão bem estão funcionando e se há algum problema. Evidentemente, isso também é importante em outras partes do mundo, como Nova Orleans e Holanda.

Ferramentas de BI e *data warehousing* tradicionalmente se concentravam em auxiliar gerentes em tomadas de decisões estratégicas e táticas através do uso de *dados históricos*. Em 2003, com o advento do *data warehousing* em tempo real, iniciou-se uma mudança em direção ao uso dessas tecnologias para decisões operacionais. Esse uso “ativo” de *data warehouses* está apenas começando a mudar o foco dessas ferramentas (veja Coffee, 2003), e IBM, Oracle, Microsoft e outros fornecedores de TI estão mudando para esses usos ativos. A Hewlett-Packard mudou para uma estratégia empresarial propícia à adaptação a fim de fornecer computação sob demanda (veja Follett, 2003; e “Infrastructure and Management Solutions”, 2003).

A tendência com vistas à produção de atualizações de dados em tempo real pelo *software* de BI para análise e tomada de decisões em tempo real está crescendo rapidamente (veja “The 2003 CIO Insight Business Intelligence Research Study”, 2003; Coffee, 2003; Patel, 2005; Raden, 2003a, 2003b; e White, 2004). Parte desse esforço envolve obter a informação certa para o pessoal operacional e tático – que geralmente lida com os aspectos a curto prazo de administrar uma empresa – para que possa usar novas ferramentas de BA e resultados atualizados para tomar decisões.

Atualmente, muitos clientes exigem dados atuais. Por isso, mais gerentes de TI estão enfrentando a questão de como conduzir uma análise de negócios em tempo real, e cada vez mais projetos de *data warehousing* em tempo real estão sendo desenvolvidos e implementados. A demanda por aplicações em tempo real continua a crescer. A proliferação de ADS e BRM (*Business Rules Management*), por exemplo, gera pressão para implementar mais processos automatizados de negócios que podem ser mais bem implementados em um ambiente de *data warehouse* em tempo real. Quando são necessários processos que exigem atualizações instantâneas para resposta de questões analíticas, é necessária uma resposta em tempo real. Conseqüentemente, os tempos de resposta para consultas, OLAP e *data mining* devem estar próximos de zero (veja Raden, 2003a).

Para conseguir uma análise de negócios em tempo real, os *data warehouses* em tempo real precisam ser atualizados com muita frequência, não apenas semanal ou mensalmente. Em 2003, a atualização diária tornou-se popular; o intervalo de tempo continua a diminuir. Além das consultas em tempo real, aplicações de BA estão sendo implementadas. Essas aplicações podem identificar instantaneamente, por exemplo, os padrões de compra de um cliente baseados na disposição da loja e recomendar alterações imediatas na localização ou na própria disposição. Outras aplicações incluem suporte à central de atendimento, detecção de fraude, gestão de rendimentos, transporte e muitas operações financeiras relacionadas.

As exigências de tempo real mudam a maneira como vemos a criação das ferramentas de bancos de dados, *data warehouses*, OLAP e *data mining*, pois elas são atualizadas enquanto

as consultas estão ativas. Por outro lado, o valor comercial substancial desse processo já foi demonstrado, então é crucial que as empresas considerem adotar esses métodos nos seus processos de negócios.

Exemplos de *softwares* de BI em tempo real baseado na Web incluem Web Intelligence (Business Objects), Supply Chain Analytics e BI Series 8 (Cognos), Live business (Data Mirror), DB2 Intelligent Miner Scoring (IMS; IBM), Analytics Delivery Platform (Informatica), PowerCenter versões padrão e avançado (Informatica), Databeacon (Databeacon Inc.), LiveSheet para Excel (KnowNow), NetIQ Corp., WebTrends, Supply Chain Intelligence Suite (SAS) e SonicMQ (Sonic Software).

Para análises críticas, consulte Havenstein (2003b) e Lindquist (2003).

Preocupações quanto aos sistemas em tempo real

Apesar dos muitos usos para os sistemas em tempo real, uma questão importante na computação em tempo real é que nem todos os dados devem ser atualizados continuamente. Com certeza, isso pode causar problemas quando os relatórios forem gerados em tempo real, pois os resultados de uma pessoa podem não combinar com os de outra. Por exemplo, uma empresa que usa o WebIntelligence (da Business Objects) observou um problema significativo com a inteligência em tempo real. Relatórios em tempo real são todos diferentes quando produzidos em períodos ligeiramente diferentes (veja Peterson, 2003). Além disso, talvez não seja necessário atualizar determinados dados continuamente.

Em muitos casos, os dados em tempo real são necessários para criação de sistemas ADS.

Suporte automatizado à decisão (ADS)

No Capítulo 1, definimos os sistemas de **suporte automatizado à decisão (ADS)** como sistemas baseados em regras que fornecem soluções para problemas gerenciais repetitivos. Os sistemas ADS também são conhecidos como sistemas de **gerenciamento de decisões empresariais (EDM)**. Os sistemas ADS estão estreitamente relacionados a BI e BA das seguintes maneiras:

- Alguns modelos de BA são usados para criar e/ou operar regras de negócios.
- As regras de negócios podem ser usadas para ativar as decisões automáticas que podem ser parte de aplicações de BI. Isso se percebe principalmente em sistemas de monitoramento e análise de desempenho nos quais os resultados do monitoramento são analisados e alguma ação é ativada pelo ADS.

Antes de observarmos como os sistemas ADS são implementados, vamos analisar as regras de negócios.

Regras de negócios

A automatização do processo de tomada de decisão geralmente é alcançada através do agrupamento da experiência do usuário de negócios em um conjunto de *regras de negócios* que são incorporadas a um mecanismo de fluxo de trabalho guiado por estas últimas (ou outra ação orientada). Essas regras de negócios podem fazer parte de um sistema especializado ou de outro sistema inteligente. Assim que a BA é chamada (p.ex., por um sistema de monitoramento de desempenho), um sinal é passado pelo mecanismo de regras para avaliação em comparação com as regras de negócios associadas. Com base nos resultados da avaliação, essas regras determinam qual ação precisa ser tomada.

White (2004) descreveu quatro tipos de regras de negócios. Normalmente, elas são encontradas em BI em tempo real e estão associadas ao *business performance management (BPM)*, conforme descrito no Capítulo 5. Os quatro tipos de regras são: regras de análise, regras de contexto, regras de exceção e regras de ação. As *regras de análise* são usadas para calcular

métricas de desempenho a partir de dados detalhados de operações comerciais, e as *regras de contexto* possibilitam que as métricas de desempenho sejam vinculadas às metas e previsões comerciais. Alguma automação básica de BPM pode ser realizada através da aplicação das *regras de exceção* às métricas e do envio de um alerta ao usuário de negócios quando uma métrica excede um limite definido na regra de exceção. A tomada de decisão pode ser mais aprimorada através da inclusão, no alerta, de um endereço na Web de um fluxo de trabalho de análise guiado. Esse fluxo identifica outras análises e relatórios que podem ser processados para investigar o problema posteriormente.

Ações automatizadas completas na hora certa podem ser realizadas através da definição do processo manual de tomada de decisão. Os usuários de negócios efetuam isso como uma série de *regras de ação* em um fluxo de trabalho. Então, esses fluxos de trabalho de ação podem ser implementados em um mecanismo de regras com o objetivo de automatizar o processo de tomada de decisão. Um ADS pode ser criado usando regras.

Características e benefícios do ADS

Os ADSs são mais compatíveis com decisões que devem ser tomadas com frequência e/ou de modo rápido, usando a informação disponível eletronicamente. Os critérios de conhecimento e decisão, assim como as regras de negócios que os expressam, devem ser altamente estruturados, e a situação do problema deve estar bem entendida. Dados e/ou conhecimento de alta qualidade sobre o campo do problema devem estar disponíveis. A aprovação de um empréstimo e a concessão de limite de crédito a um cliente são exemplos típicos deste tipo de decisão.

Usar o ADS pode motivar tomadas de decisão precisas e ágeis, pois ele oferece os seguintes recursos:

- Criação rápida de aplicações baseadas em regras para automatizar ou guiar os tomadores de decisão e implementação das aplicações em quase todos os ambientes operacionais.
- Introdução de análise preditiva nas aplicações baseadas em regras, aumentando seu poder e valor.
- Fornecimento de serviços de decisão aos sistemas legados, ampliando seus recursos enquanto minimizam o risco técnico.
- Combinação flexível de regras de negócios, modelos preditivos e estratégias de otimização em aplicações de gestão de decisão modernas.
- Aceleração da percepção de aprendizagem proveniente de critérios de decisão em criação, execução e refinamento da estratégia.

Aplicações de ADS

Davenport e Harris (2005) forneceram vários exemplos de sistemas ADS bem-sucedidos em diversos setores. Alguns exemplos típicos são:

- *Configuração de produto ou serviço*. Clientes são autorizados a personalizar um produto (ou serviço), como por exemplo, um PC. Então, o ADS configura o produto (ou serviço) final mais adequado e seu custo (considerando a rentabilidade para o fabricante).
- *Otimização do lucro (preço)*. Empresas aéreas usam extensivamente aplicações de tomada de decisão para definir preços com base na disponibilidade de assentos e hora ou dia da compra, também conhecido como *gerenciamento de rendimento* ou *gerenciamento de lucro*. Outras empresas de fabricação, logística e transporte usam essas aplicações para aprimorar a eficiência operacional. Um método, conhecido como otimização de preço ou precificação inteligente (veja Fleischmann et al., 2004), está sendo usado de forma experimental com a precificação variável (p.ex., aluguel imobiliário). Para exemplos de precificação de financiamento imobiliário, veja Barret e Carr (2005).

- *Decisões de encaminhamento ou segmentação.* Melhorias significativas na produtividade foram feitas por empresas que criaram filtros automáticos para classificação de casos e operações. Exemplos incluem o setor de seguros, onde reclamações nítidas são tratadas como “prioridade”, e salas de emergência, que administram uma variedade e volume de pacientes.
- *Conformidades corporativa e regulatória.* Decisões rotineiras sobre normas são técnicas e consomem tempo; entretanto, as regras devem ser atuais e precisam ser aplicadas consistentemente. Por exemplo, no setor de financiamento imobiliário, os credores devem classificar e processar empréstimos que estejam em conformidade com os regulamentos governamentais e também com as exigências dos credores. Se o processo for concluído de forma eficiente, a economia de custos é significativa.
- *Deteção de fraude.* A Receita Federal e a Comissão de Valores Mobiliários dos EUA usam a triagem automatizada para identificar fraudes. Bancos e outras empresas de financiamento devem aprimorar seus sistemas devido à lei Sarbanes-Oxley e outras. Para exemplos, veja Barret e Carr (2005).
- *Previsão dinâmica.* Maior previsão automatizada de demanda por parte dos fabricantes permite às empresas alinharem as previsões do cliente (ou seja, controle de níveis de estoque) com seus planos de fabricação e venda.
- *Controle operacional.* Alguns sistemas ADS são programados para perceber mudanças no ambiente físico e responder com base em regras ou algoritmos (p.ex., temperatura afetando as necessidades de fornecimento de energia, semáforos em cruzamentos baseados em volume de tráfego em tempo real gravado por câmeras).

Implementação do ADS

Empresas de *software* oferecem os seguintes componentes para ADS:

- *Mecanismo de regras.* Empresas como a Exsys (exsys.com), a ILOG (ilog.com) com o BRMS, a Fair Isaac (fairisaac.com) com o Blaze Advisor para análise de risco e o Falcon para análise de fraude, a ARulesXL (arulesxl.com) e a Pegasystems (pega.com) oferecem processamento de regras de negócios.
- *Algoritmos matemáticos e estatísticos.* Empresas como a SAS e a SPSS oferecem fórmulas para descoberta de soluções mais adequadas (p.ex., preço, nível de estoque), acompanhamento de análises de tendência e assim por diante. (Veja a Seção 3.6.) A SAS, por exemplo, oferece precificação e otimização de produto (veja sas.com/success/autozone.html).
- *Pacotes específicos para o setor.* Dúzias de empresas fornecem pacotes de *software* para setores específicos (p.ex., LendingTree para o setor de financiamento imobiliário e ao cliente, CSC Continuum para o setor de seguros).
- *Sistemas empresariais.* Empresas como a SAP e a Oracle oferecem aplicações que automatizam, reúnem e direcionam fluxos de informação e processamento de operações em organizações complexas que usam tecnologia automatizada de decisão para funções específicas (p.ex., SCM).
- *Aplicações de fluxo de trabalho.* Após um mecanismo de regras tomar uma decisão, o sistema de fluxo de trabalho executa processos de negócios repletos de informações pelas etapas exigidas. Entre os principais fornecedores desses sistemas estão a Documentum e a FileNet Corp.

Inteligência competitiva

Informações sobre o concorrente podem significar a diferença entre ganhar ou perder uma batalha nos negócios. Muitas empresas monitoram continuamente as atividades de seus concorrentes a fim de adquirir *inteligência competitiva*. Essa junção de informações impulsiona o desempenho comercial ao aumentar o conhecimento de mercado, melhorar a gestão de conhecimento e elevar a

qualidade do planejamento estratégico. Considere, por exemplo, os seguintes usos de inteligência competitiva baseada em computador, citados por Comcowich (2002):

- Uma empresa de produtos esportivos descobriu, com meses de antecedência, um grupo de ativistas planejando uma manifestação e um boicote, permitindo a implantação de uma contra-estratégia.
- Alguns dias após o lançamento, uma empresa de *software* constatou um descontentamento com recursos específicos do produto, permitindo aos técnicos criar um *patch* que consertou o problema em dias no lugar dos meses exigidos normalmente para obter retorno do cliente e implementar as correções de *software*.
- Uma empresa de embalagens conseguiu determinar a localização, o tamanho e a capacidade de produção de uma nova planta que será construída por um concorrente. A informação bem-protegida foi descoberta por um serviço automatizado de monitoramento em concessões de construção no *website* da cidade onde a nova planta será construída.
- Uma empresa de telecomunicação descobriu uma estratégia legislativa de um concorrente, permitindo a obtenção de vantagem em uma intermediação estado a estado. (O interessante é que a estratégia estava postada no próprio *site* do concorrente.)
- Uma equipe criativa que iniciava o desenvolvimento de um novo *videogame* usou a Internet para identificar características de vanguarda do produto que eram de preferência dos jogadores. Uma pesquisa intensa descobriu três “tem que ter” fundamentais que não foram identificados nos grupos focais nem incluídos na especificação original do projeto.

A inteligência competitiva pode ser facilitada com tecnologias como reconhecimento óptico de caracteres, identificação por radiofrequência (RFID), agentes inteligentes e, em especial, a Internet. A Internet está se tornando a ferramenta mais importante de uma empresa no suporte à inteligência competitiva. A visibilidade da informação que um concorrente coloca na Internet e a capacidade das ferramentas baseadas na Web de interrogar *websites* sobre informações de preços, produtos, serviços e abordagens de *marketing* tem gerado grande interesse corporativo nessas atividades de agrupamento de inteligência. Por exemplo, a livraria *online* especializada FatBrain (agora parte da **barnesandnoble.com**) usou a empresa de “e-spionagem” Rivalwatch para acompanhar a concorrência no mercado de livros especializados e educacionais da Fatbrain. Ao controlar preços das empresas rivais como a Amazon.com, a Fatbrain pôde oferecer preços competitivos sem entregar suas margens de lucro quando não era necessário.

Pawar e Sharda (1997) propuseram uma estrutura na qual os recursos da Internet são mostrados a fim de fornecer informações para decisões estratégicas. De acordo com essa estrutura, as informações externas exigidas e os métodos de aquisição de informação podem receber suporte de ferramentas da Internet para comunicação, pesquisa, busca e recuperação de informações. Pawar e Sharda enfatizam a capacidade de pesquisa das diversas ferramentas da Internet. Ao usar essas ferramentas, uma empresa pode implementar estratégias específicas de pesquisa, conforme ilustrado no Arquivo *online* W3.5.

Entretanto, apenas colher informações sobre um concorrente não é o suficiente. Analisá-las e interpretá-las é tão importante quanto coletá-las. Para essas tarefas, podemos usar ferramentas de BI que variam de *agentes inteligentes* (ou seja, ferramentas de *software* que permitem a automação de tarefa que exigem inteligência) à *data mining* (ou seja, pesquisar em grandes bancos de dados relações entre *bits* de dados, usando ferramentas lógicas especializadas; veja Capítulo 4). A J.P. Morgan Chase (Nova Iorque), por exemplo, usa *data mining* para controlar diversas fontes de informação. O objetivo da empresa é determinar um possível impacto da informação para o banco, setor e os clientes.

Outro aspecto, mais perverso, da inteligência competitiva é a *espionagem industrial*. Os espionagens corporativos, que realmente existem em alguns setores, procuram planos de *marketing*, análise de custos, produtos/serviços propostos e planos estratégicos confidenciais. A espionagem industrial é considerada antiética e, em geral, ilegal. Um tipo de espionagem industrial é o roubo de computadores portáteis em aeroportos, hotéis e conferências. Muitos dos ladrões estão mais interessados nas informações armazenadas nos computadores do que nos computadores em si. Proteger-se contra essas atividades é uma parte importante da manutenção da vantagem competitiva.

A revista *Baseline* dedicou sua edição de 15 de dezembro de 2005 às ameaças contra a segurança, incluindo ataques de concorrentes (p.ex., Gage, 2005). Porém, segundo a SCIP (*Society of Competitive Intelligence Professionals*; scip.org), a inteligência competitiva não inclui espionagem. *Espionagem* é o uso de meios ilícitos para obter informações. Por outro lado, a *inteligência competitiva* usa meios legais e éticos para obter e analisar informações. Muitas empresas ajudam na inteligência competitiva (p.ex., fuld.com). Para recursos abrangentes em BI, acesse b-eye-network.com.

Questões de revisão da Seção 3.9

1. Defina *BI em tempo real*.
2. Liste os benefícios da BI em tempo real.
3. Quais são as tecnologias necessárias para oferecer BI em tempo real?
4. Liste algumas preocupações relacionadas aos sistemas em tempo real.
5. Defina *sistema ADS* e descreva os benefícios do ADS.
6. Descreva gerenciamento de lucro.
7. Liste as principais categorias de ADS.
8. Descreva inteligência competitiva.
9. Como a Internet pode facilitar a inteligência competitiva?

3.10

ANÁLISE DE NEGÓCIOS E A WEB: INTELIGÊNCIA DA WEB E ANÁLISE DA WEB

As atividades de BI – desde a aquisição de dados até a armazenagem, BA e *mining* – são realizadas, na maioria das vezes, usando ferramentas da Web. Usuários com navegadores podem acessar um sistema, fazer pesquisas, obter relatórios, etc., tudo em tempo real. Isso é feito através de *intranets* para executivos com informações privilegiadas e via Internet ou *extranets* para pessoas de fora.

Uso da Web em BA

A maioria das aplicações de BA está relacionada à Web. Fornecedores de *software*, por exemplo, oferecem ferramentas da Web que conectam um *data warehouse* a sistemas de pedidos e catalogação para comércio eletrônico. Um exemplo é o RETSbond, um produto da HotScript (hotscripts.com). O pacote de ferramentas para comércio eletrônico da Hitachi combina atividades de comércio eletrônico, como gerenciamento de catálogos, aplicações de pagamento, personalização em massa e gerenciamento de pedidos com *data warehouses* (ou *data marts*) e sistemas ERP.

Fornecedores de *data warehousing*, análise e suporte à decisão estão integrando seus produtos às tecnologias da Web e ao comércio eletrônico e também estão criando novas tecnologias para o mesmo objetivo. Alguns exemplos são o Web Intelligence da Business Objects e o produto Appsource “programado para OLAP” da Hyperion, que integra OLAP com ferramentas da Web. O Decision Edge da IBM é outra ferramenta que oferece recursos OLAP em uma *intranet* de qualquer lugar de uma empresa, usando navegadores, mecanismos de pesquisa e outras tecnologias da Web. BA também pode ser facilitada através do uso de produtos, como por exemplo, o Financial Analyzer e o Sales Analyzer da Oracle, e o BI/Web e o BI/Broker da Hummingbird.

Análise da Web

Análise da Web é a aplicação de atividades de BA a processos baseados na Web, incluindo comércio eletrônico. Esse termo é usado para descrever a aplicação de BA aos *websites*. Na realidade, as ferramentas e os métodos são altamente visuais. Schlegel (2003) descreveu as bases da análise da Web e forneceu uma sugestão de arquitetura para a *análise de clickstream* (ou seja, uma aplicação da análise da Web). Para uma descrição abrangente, consulte Peterson (2005). Para uma visão geral sobre análise, métodos e produtos de *websites*, consulte Rapoza (2004).

Análises de acesso em websites

Análise de acesso em websites refere-se à análise de dados que acontece dentro do ambiente da Web; os dados, conhecidos como **dados de acesso à Web**, fornecem o caminho das atividades do usuário e mostram os padrões de navegação do usuário: quais *sites* são visitados, quais páginas são acessadas, quanto tempo é gasto em um *site*, etc. Ao analisar e interpretar esses dados, uma empresa pode, por exemplo, descobrir a eficácia de promoções e determinar quais produtos e anúncios chamam mais atenção.

À medida que as operações de acesso à Web aumentam, a quantidade de dados a ser processada aumenta exponencialmente, e questões de escalabilidade tornam-se fundamentais para análise da Web. Werner e Abramson (2001) descreveram um método (baseado em classificação e agrupamento) para processar 1 bilhão de registros por dia para um *data warehouse* na Web. Veja também Ruber (2001) para informações sobre análise de acesso à Web.

Segundo Edelstein (2001), a realização de uma análise de acesso à Web exige o seguinte:

- Dados brutos de acesso à Web devem ser obtidos de servidores múltiplos.
- Dados particulares de clientes geralmente são misturados com outros dados sobre páginas servidas, *hosts*, páginas de referência e tipos de navegador.
- Uma única solicitação de página pode gerar múltiplas entradas em registros de servidor.
- Criar uma sessão de visualizações de página baseada em uma sequência de registros envolve a limpeza de muitos dados a fim de eliminar dados supérfluos.
- Identificar sessões em um fluxo de dados é complexa. Exige *cookies* ou outros números de identificação da sessão em URLs.
- Servidores *proxy* (ou seja, onde as solicitações do cliente não vêm do servidor doméstico) confundem a identidade de uma sessão e o porque foi encerrada.
- A preparação de dados pode consumir 80% dos recursos de um projeto.

Suporte do fornecedor à análise da Web

A maior parte dos fornecedores oferece suporte extensivo na Web para seus produtos de BI. A seguir seguem alguns exemplos ilustrativos:

- A *Business Objects* oferece uma solução cliente completo capacitada para Web para consulta e análise chamada Web Intelligence. Ela permite aos usuários de negócios controlar e gerenciar, de maneira fácil, informações armazenadas em diversas fontes de dados dentro e fora da empresa, de forma integrada, a partir de um computador *desktop*.
- A *Cognos* oferece uma arquitetura de Web *services* completa no seu produto Cognos 8. Ele permite um desenvolvimento de *software* mais fácil.
- A *Informatica* tem se concentrado muito no uso da Web para permitir às empresas controlar o desempenho comercial. Ao usar a plataforma de BI Power Center 8, as empresas reúnem métricas de desempenho comercial através de sistemas de voz, da Web e de transmissões

sem fio. A Analytics Delivery Platform da Informatica é um sistema baseado na Internet que fornece resultados de desempenho do negócio em tempo real.

- A ferramenta de BI *WebTrends* se concentra na análise em tempo real do tráfego na Web e das operações *online*. Ela permite às empresas controlar as tendências de compra do cliente, o lucro e a eficácia de campanhas publicitárias ou promoções de vendas por meio de milhões de visitas aos *sites* diariamente. O Site59.com Inc., um site de viagem especializado em pacotes de partida de última hora, descobriu com a análise ao vivo da *WebTrends* que os visitantes não conseguiam achar facilmente todos os pacotes de viagem disponíveis no *site*. A análise indicou como aprimorar e melhorar a estrutura. Desde então, o Site59.com constatou um aumento no número de visitantes e na proporção daqueles que fazem compras *online* (veja Pallatto, 2002).
- O *Google* oferece de graça seu *Google Analytics* para pequenas empresas. É um *software* para análise de acesso de registros na Web.
- A *ADVIZOR Solutions* oferece uma biblioteca interativa de gráficos que suporta necessidades de exibição comercial, análise integrada (p.ex., Visual Discovery, análise preditiva), análise sob demanda de diversas fontes de dados em tempo real, e auto-serviço apontar e clicar usando o Visual Discovery. Para mais detalhes, acesse advizorsolutions.com/solutions.htm.
- O KnowledgeStudio da *Angoss* é outro exemplo de ferramenta de análise da Web/*Web mining*. Veja a Figura 3.7 para uma amostra de captura de tela.

A utilização da análise de acesso na Web é demonstrada no Caso de aplicação 3.6.

CASO DE APLICAÇÃO 3.6

Análise da Web aperfeiçoa o desempenho do comércio eletrônico

As vendas de comércio eletrônico estão crescendo rapidamente, estimuladas pelo crescimento de novas categorias de produto e habilidade varejista de *marketing* digital, análise da Web e *marketing* multicanal.

Comerciantes virtuais, ansiosos para melhorar o retorno de investimento dos seus *websites*, devem se informar, em tempo real, sobre as atividades dos seus visitantes. Isso vai muito além de realizar análise de acesso à Web padrão e obter relatórios de operação com ferramentas distintas.

É fundamental entender o comportamento *online* do cliente para determinar como e o que anunciar e comercializar para ele. Entender e usar adequadamente as métricas operacionais de um *site* de comércio eletrônico pode fazer um negócio crescer ou falir. Por exemplo, promoções especiais de produtos podem ser colocadas *online* em questão de dias, em oposição aos meses exigidos para caras revisões de catálogo e malas diretas por todo o país.

Análise da Web pode impulsionar o lucro líquido. A diretora de Internet da Newport News (newport-news.com), Yun-Hui Chong, disse que os dados que sua empresa coleta nas atividades dos seus aproximadamente 2 milhões de visitantes mensais na Web per-

mite avaliar o retorno do investimento (ROI) de todas as suas campanhas de *marketing online*. Com base nisso, a empresa otimiza *banners* e a apresentação de produtos. Ela também usa as informações para fazer análise de acesso à Web a fim de entender como os clientes estão reagindo às mudanças no *site*. Tem se tornado especialmente importante identificar clientes que abandonam o *site* ou navegam em apenas determinadas categorias. A empresa envia a esses clientes por *e-mail* promoções personalizadas das suas categorias preferidas. Desde que a empresa começou a fazer isso, houve um aumento significativo na conversão de *e-mail* e no rendimento por *e-mail* enviado. Navegadores e abandonadores direcionados via *e-mail* em três categorias de produto que tinham as piores taxas de conversão resultaram em uma melhora superior a 6 vezes no aumento da receita por *e-mail* enviado, enquanto o custo por pedido caiu cerca de 83%. A análise da Web nitidamente vale a pena!

Fontes: compilado de P. Ruber, "Analytics Improve Merchandising", *Internetworld*, June 2003, and V. Rhodes, "SAS Improves Evaluation of Campaign Segments for Newport News", *DM Review Magazine*, March, 2002.

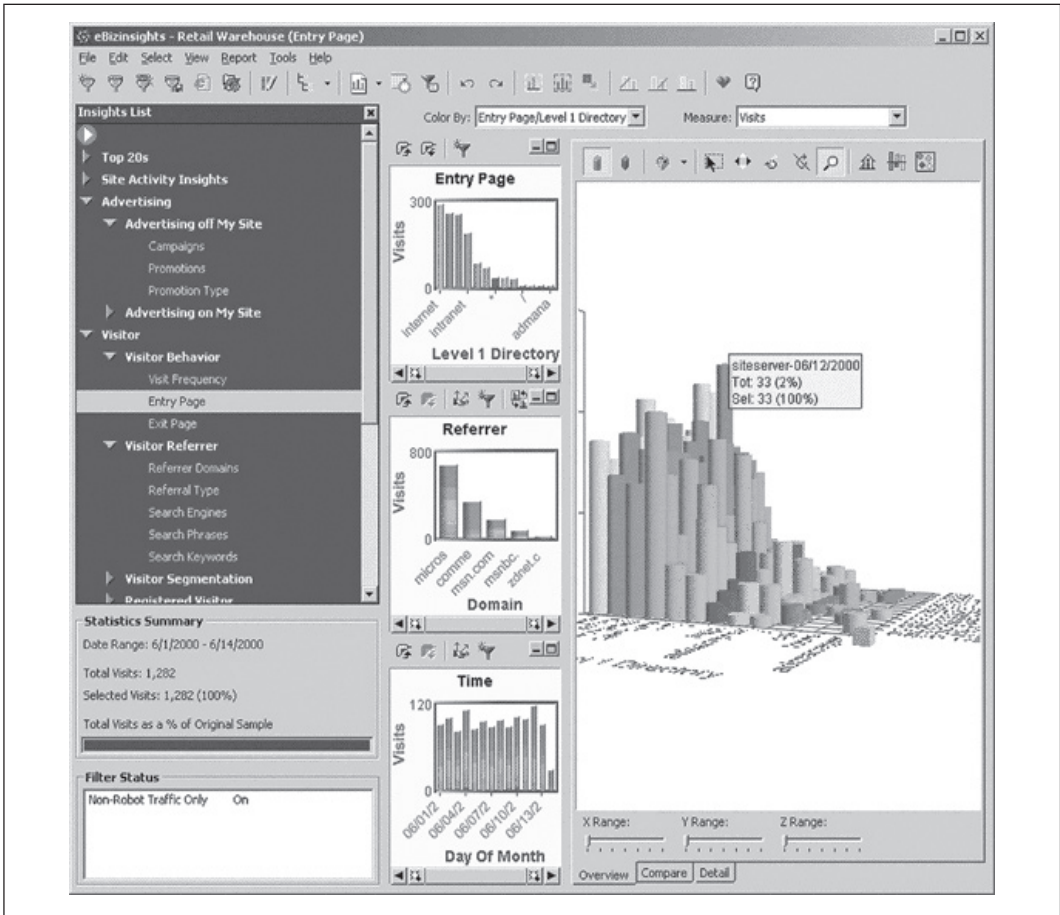


Figura 3.7 Captura de tela da análise visual de desempenho de *website* do eBizInsights.

Fonte: cortesia da Visual Insights.

Questões de revisão da Seção 3.10

1. Descreva o uso da Web para aprimorar as atividades de BI.
2. Defina *análise da Web*.
3. Descreva as aplicações da análise da Web.
4. Defina *análise de acesso na Web*.
5. Como os dados de acesso na Web são usados?

3.11

USO, BENEFÍCIOS E SUCESSO DA ANÁLISE DE NEGÓCIOS

O uso de BA está se expandindo rapidamente, e a maioria das pequenas e médias empresas a usam com êxito. A seguir, examinaremos algumas questões de uso e implementação.

Uso de BA

Quase todos os gerentes e executivos podem usar algum sistema de BA, mas alguns acham as ferramentas muito complicadas de usar. Algumas vezes, os gerentes não são treinados de forma

adequada. Distribuir ferramentas analíticas por toda uma empresa é um grande desafio: muitos negócios exigem um grande percentual da empresa para potencializar a análise, mas a maioria dos desafios relacionados à adoção da tecnologia envolve cultura, pessoas e processos (veja Bonde e Kuckuk, 2004). Uma questão fundamental é alinhar aplicações de BA às necessidades comerciais. Se o sistema não fornecer informações úteis, é considerado inútil. Consulte os *Insights* de tecnologia 3.7 para detalhes de um estudo sobre como os executivos usam ferramentas de BI e BA. Dados adicionais e casos de sucesso podem ser encontrados em Thompson (2004).

Sucesso e aplicação de BA

De acordo com um relatório do IDC do segundo semestre de 2002, empresas que implementaram e usaram com êxito aplicações analíticas obtiveram retornos que variam de 17% a mais de 2.000%, com um ROI médio de 122% (veja Morris et al., 2002). Mesmo assim, mais da metade de todos os projetos de BI falham. Assim como *data warehousing*, as atividades de BI devem ser consideradas não apenas outro conjunto de projetos de TI, mas sim uma constante envolvendo estratégia, visão e arquitetura que busca de forma contínua alinhar as operações e a direção de uma empresa com suas metas comerciais estratégicas. Para ver uma história de sucesso, consulte o Arquivo *online* W3.6.

Em empresas, as ferramentas de BI têm sido usadas para identificar crimes do colarinho branco. Elas podem ser usadas para identificar notas fiscais superfaturadas, desfalques, roubo da identidade do cliente e crimes semelhantes. Segundo Dorrington (2003), a estimativa do total de fraudes no Reino Unido é de quase US\$30 bilhões. Fraudes cometidas por funcionários causam perdas médias de US\$60.000, e fraudes cometidas por gerentes ou executivos causam perdas mé-

INSIGHTS DE TECNOLOGIA 3.7

Os sistemas de BI tornam as empresas mais inteligentes?

Mais de 570 executivos de TI responderam à pesquisa sobre *business intelligence* da *CIO Insight*. A revista descobriu alguns fatos interessantes sobre a situação de BI e BA em 2003 (e que continua atual):

- O mais notável, o uso de tecnologias de BI é relativamente alto, e está crescendo.
- Geralmente, grandes empresas estão mais propensas a usar BI do que pequenas empresas.
- Em 2002, empresas bem-sucedidas gastavam quase 50% a mais em tecnologia de BI do que empresas malsucedidas. A BI parece ser necessária (mas não suficiente) para o sucesso.
- Na prática, o governo usa toda ferramenta de BA disponível comercialmente em uma porcentagem significativamente maior do que qualquer outro setor da economia.
- Na inteligência competitiva, as tecnologias usadas para coletar, reunir, analisar e relatar, junto com a porcentagem de resposta em parênteses, são: ferramentas para geração de relatórios (82,1%), transmissão automatizada de dados/informações (79%), *intranets*/portais (70,4%), *data warehousing* (69,8%), gerenciamento de conteúdo (63%), *software* para visualização de dados (41,4%), *software* para fluxo de trabalho (41,4%) e coleta automática de metadados (p.ex., agentes inteligentes) (38,9%).
- Apenas 49% das empresas menos bem-sucedidas estão satisfeitas com suas iniciativas de inteligência competitiva.
- Cerca de 88% das empresas confiam na precisão das informações sobre os clientes que reúnem usando ferramentas de BI.
- A insatisfação com BI normalmente resulta da dificuldade em distribuir os resultados.
- Os CIOs desejam mudar as empresas para que sejam empresas em tempo real que usam BI.

Fonte: compilado de “The 2003 CIO Insight Business Intelligence Research Study: Are Your BI Systems Making You Smarter?” *CIO Insight*, May 23, 2003, and *CIOs Rate SAS Tops for Increasing ROI, Delivering Business Value and Reliability*, February 9, 2004, sas.com/news/preleases/020904/news1.html (acessado em fevereiro de 2006).

dias de US\$ 250.000. Quando gerentes e executivos trabalham em conjunto, a perda média sobe para US\$ 500.000.

Se todos os sistemas de dados internos estivessem integrados a um *data warehouse* para análise de fraudes, os dados internos poderiam ser comparados a dados externos relacionados à fraude para que padrões e irregularidades tornem-se identificáveis de forma mais rápida. Atividades suspeitas podem ser isoladas, avaliadas e rastreadas.

A Williams-Sonoma Corp. economizou milhões com *marketing* direcionado e *branding* multicanal ao usar o *software* de *data mining* da SAS, o Enterprise Miner, junto com um pacote de aplicações CRM da SAS. O novo sistema de *marketing* formula e explora dados do cliente de mais de 30 milhões de famílias a fim de ajudar o varejista a criar uma experiência de compra personalizada e coesa em diversos canais e marcas. Veja Bolen (2003) para mais detalhes. Callaghan (2003a) descreveu como o Predictive Web Analytics da SPSS e o SAS Web podem ser usados para prever o comportamento do cliente na Web e desenvolver modelos (*clusters*) de segmentação de cliente que levam a um melhor desempenho comercial. Frequentemente, varejistas usam ferramentas de BI, como mostrado no Caso de aplicação 3.7.

CASO DE APLICAÇÃO 3.7

Varejistas tornam estável o progresso do BI

A maioria dos varejistas convencionais fica para trás de outros setores no uso de BI. Exceções notáveis incluem Wal-Mart Stores Inc. e Sears. Outros varejistas continuam a fazer progressos impressionantes. Seguem alguns exemplos:

- Hudson's Bay Co. fez 333 anos em maio de 2003. Apesar da idade, a Hudson's Bay atualizou seus sistemas de informação a fim de dar aos executivos, gerentes de loja e principais fornecedores métodos para analisar resmas de dados de vendas e clientes. O desafio enfrentado pela empresa é determinar como transformar os dados em informações úteis. A empresa usa dois *data warehouses* e ferramentas de BI da Teradata para controlar as vendas e tomar decisões sobre estoque e venda de produto.
- Na Harry Rosen Inc., uma cadeia com 17 lojas de vestuário masculino, os executivos usam as ferramentas de análise de dados da Cognos Inc. integradas ao sistema de *merchandising*. Mais de uma dúzia de relatórios de venda e estoque para análise de vendas ajuda a empresa a identificar tendência de vendas, gerenciar estoque e melhorar as margens de lucro bruto.
- A Reno-Depot, uma loja varejista canadense para reformas domésticas com mais de 20 lojas, usa informação de tráfego de clientes, dados de vendas dos pontos de vendas, etc. para desenvolver, prever e criar cronogramas

de trabalho adequados. O *software* realiza uma análise das situações de trabalho existentes, criando dinamicamente turnos com as tarefas do trabalhador.

- A Coldwater Creek gera relatórios através da integração de dados produzidos em aplicações muito diferentes usando *data warehousing* e o Microsoft SQL Server 2000. O tempo de preparação de relatórios foi reduzido significativamente.
- Ao usar ferramentas de BI e análise da Business Objects, a TruServ Corp. (controladora da True Value Hardware e da Taylor Rental) reduziu seu estoque de “zona vermelha” (ou seja, produtos que não foram vendidos em meio ano) em US\$50 milhões em dois anos através da análise das reservas de produto. O sistema também identificou produtos “encalhados” nos 14 centros de distribuição da empresa que poderiam vender melhor em outras partes do país.

Outros varejistas estão buscando maneiras similares de obter uma vantagem competitiva. Colocar o produto certo no local certo, na hora certa e com o preço certo é uma das metas dos varejistas. Fazer isso corretamente determina quem tem sucesso e quem falha.

Varejistas tradicionais estão aprendendo com os varejistas virtuais como realizar investigações analíticas sobre desempenho do cliente. A J. Crew Group e a Nordstrom Inc., por exemplo, usam o DigiMine

para analisar as vendas *online*. A Nordstrom tinha uma situação na qual compradores virtuais estavam a procura de *piercings* de umbigo exatamente iguais ao que uma modelo usou em um anúncio. A empresa conseguiu obter rapidamente os *piercings* para suas lojas e seus clientes virtuais, mesmo não tendo o produto antecipadamente.

Um grande número de varejistas está recorrendo a programas automáticos para determinação de preços.

Fontes: compilado de R. Whiting, “Business-Intelligence Buy-in”, *InformationWeek*, May 12, 2003; “Hammacher Schlemmer & Company, Inc.”, *stores.com* (acessado em abril de 2006); and “A Breath of Fresh Air: Coldwater Creek Freshens Up Its Reporting”. *Retail Technology Quarterly*, July 2005.

Novas formas de BI continuam a surgir. Sistemas de gestão de desempenho (PMS), também conhecidos como gestão de desempenho corporativo (COM; veja o Capítulo 5), são uma dessas novas formas. Eles são ferramentas de BI que fornecem indicadores e outras informações relevantes que os tomadores de decisão podem usar para determinar seus níveis de sucesso na obtenção das metas.

Por que projetos de BI/BA falham

Empresas devem entender e enfrentar muitos desafios críticos para atingir o sucesso em BI. Segundo Atre (2003), existem 10 principais razões para os projetos de BI falharem:

1. Falha ao reconhecer os projetos de BI como iniciativas comerciais interorganizacionais e ao entender que, como tal, diferem de soluções típicas autônomas
2. Patrocinadores comerciais fracos ou não comprometidos
3. Representantes comerciais de áreas funcionais inacessíveis ou relutantes
4. Falta de pessoal capacitado (ou disponível) ou utilização de pessoal abaixo do ideal
5. Falta de conceito de liberação de *software* (ou seja, sem método de desenvolvimento iterativo)
6. Falta de estrutura de paralisação de trabalho (ou seja, sem metodologia)
7. Falta de análise de negócios ou atividades de padronização
8. Falta de avaliação do impacto negativo de “dados sujos” na rentabilidade comercial
9. Falta de compreensão da necessidade e do uso de metadados
10. Muita dependência de métodos e ferramentas bastante diferentes

Desenvolvimento de sistema e a necessidade de integração

Desenvolver uma aplicação de BI eficiente para suporte à decisão pode ser bastante complexo; para metodologia e orientação veja Moss e Atre (2004). A integração, quer de aplicações, fontes de dados ou mesmo ambiente de desenvolvimento, é um FCS para BI. Por essa razão, a maioria dos fornecedores de BI (com destaque para Oracle, BusinessObjects, MicroStrategy, IBM e Microsoft) oferece acervos de aplicações altamente integrados, incluindo conexão ao ERP e CRM. Ho et al. (2004) forneceram um exemplo de integração OLAP que usa *data warehouses* e redes neurais (veja o Capítulo 6). No exemplo, o resultado de OLAP em um sistema de fluxo de trabalho logístico foi analisado pela rede neural. A maioria dos fornecedores de BI está preparada para integração de aplicações (veja Callaghan, 2005), geralmente habilitada para Web (veja *businessobjects.com*).

Questões de revisão da Seção 3.11

1. Liste os principais benefícios da BA.
2. Liste três fatores relacionados à implementação de BI.
3. Identifique algumas ferramentas de BI.
4. Liste 5 razões por que os projetos de BI falham.

Destaques do capítulo

- BA envolve muitos métodos que podem ser organizados de diferentes formas. Esses métodos podem ser classificados como descoberta de informação, suporte à decisão e ferramentas de CPM. BA geralmente é realizada em dados no *data warehouse*.
- Uma outra classificação de BA inclui relatórios, consultas, análise avançada, multidimensionalidade, indicadores e *dashboards*.
- OLAP é um conjunto de ferramentas para análise de dados oportuna. É extremamente importante nas aplicações de MSS/BI.
- EIS/ESS são os predecessores da BI e BA. Seus principais recursos são *drill-down*, KPI, relatórios de status, análise de tendência, FCS e visões parciais na Web.
- OLAP é um termo guarda-chuva para as ferramentas de BI desenvolvidas nos anos 90. Elas incluem principalmente geração de relatórios, consultas, análise estatística simples e visualização. Os tipos mais importantes de OLAP são o MOLAP e o ROLAP.
- Os relatórios podem ser de rotina ou sob demanda (*ad hoc*). Existem dúzias de tipos de relatórios para áreas funcionais diferentes.
- A multidimensionalidade de dados permite às pessoas visualizar dados rapidamente em dimensões diferentes, mesmo se os dados estiverem localizados em diferentes arquivos e bancos de dados. É o núcleo do OLAP.
- Os relatórios avançados incluem alerta e *dashboards*. Eles podem ser consultados de maneira rápida e fácil.
- As consultas *ad hoc* podem ser realizadas de maneiras inteligentes, usando sistemas inteligentes. SQL é a principal ferramenta de consulta.
- Os cubos de dados permitem geração rápida de consultas e relatórios pelos usuários finais.
- BA avançada inclui análise preditiva e *data mining*.
- BI está mudando em direção aos recursos em tempo real.
- A visualização de dados é um recurso de BI importante.
- A análise preditiva usa diferentes algoritmos para prever resultados e relações entre variáveis, e também para identificar padrões de dados.
- A visualização de planilhas possibilita apresentações imediatas de dados complexos em uma única imagem.
- GIS apresenta dados geográficos de referência como mapas digitais. Eles podem dar suporte às tomadas de decisão em muitas aplicações que são relacionadas a localizações.
- GIS pode ser integrado ao GPS para criar muitas aplicações, em especial aquelas relacionadas a transporte e I-commerce.
- Muitas decisões comerciais precisam ser tomadas em tempo real (ou muito perto disso). Dar suporte a essas decisões com BA exige um *data warehouse* em tempo real e recursos especiais de BA.
- ADS usa regras de negócios, que geralmente são executadas com sistemas inteligentes, para dar suporte a um grande número de decisões repetitivas, como otimização de preços e configuração de produtos.
- A inteligência competitiva pode ser realizada na Internet através do uso de ferramentas de BA.
- Um número considerável de aplicações de BA é realizado na Web.
- A inteligência da Web (ou análise da Web) refere-se à análise de dados da Web (conhecidos como dados de acesso na Web). Essas análises são úteis para pesquisa de mercado e inteligência competitiva.
- Quase todas as empresas, de médias a grandes, usam algum tipo de BA, com ou sem um *data warehouse*, para economia significativa de custos.
- Para garantir o sucesso de projetos de BI e BA, é importante identificar fatores de sucesso (p.ex., participação do usuário) e cultivar esses fatores.

Termos-chave

- | | | |
|-------------------------------------|---|--|
| • análise | • cubo de dados | • processamento analítico <i>online</i> (OLAP) |
| • análise da Web | • dados de acesso na Web | • relatório de status |
| • análise de acesso na Web | • <i>drill-down</i> | • sistemas de informação geográfica (GIS) |
| • análise de negócios (BA) | • gestão de decisão empresarial (EDM) | • sistemas de posicionamento global (GPS) |
| • análise preditiva | • linguagem de consulta estruturada (SQL) | • suporte automatizado à decisão (ADS) |
| • banco de dados multidimensional | • multidimensionalidade | • visualização de dados |
| • <i>business intelligence</i> (BI) | • OLAP multidimensional (MOLAP) | |
| • consulta <i>ad hoc</i> | • OLAP relacional (ROLAP) | |
| • cubo | | |

QUESTÕES PARA DISCUSSÃO

1. Relacione *data warehouse* ao OLAP e à visualização de dados.
2. Compare OLTP ao OLAP.
3. Explique a relação entre OLAP e *data mining*.
4. Descreva multidimensionalidade e explique seus possíveis benefícios para MSS.
5. Discuta os benefícios estratégicos da BA.
6. Descreva os conceitos básicos da inteligência da Web e análise da Web.
7. Por que os fornecedores que oferecem ferramentas de ERP (p.ex., SAP, Oracle) também oferecem ferramentas de BA?
8. Compare *data mining* e análise preditiva, e discuta por que alguns pensam que eles são semelhantes enquanto outros pensam o oposto.
9. A BI substituirá o analista de negócios? Discuta. (Sugestão: veja McKnight, 2005).
10. As ferramentas de ADS substituirão o analista de negócios?
11. Discuta os benefícios do GIS como suporte à visualização para tomada de decisão.
12. Diferencie análise preditiva de *data mining*. O que eles têm em comum?
13. Relacione análise competitiva à BI.
14. Discuta como ADS pode dar suporte aos funcionários da linha de frente (p.ex., aqueles que fornecem atendimento ao cliente).
15. Por que BA em tempo real está se tornando fundamental?
16. Relacione análise avançada ao ERP e à SCM.
17. Discuta a relação entre visualização e Excel.

Exercícios

Exercícios da Teradata University e outros exercícios interativos

1. Acesse teradatastudentnetwork.com. Descreva desenvolvimentos recentes no campo.
 - a. Encontre o seminário na Web sobre visualização de informação. Responda as seguintes questões:
 - i. Quais são os recursos dos produtos da Tableau Software?
 - ii. Compare as duas apresentações e cite semelhanças e diferenças entre elas.
 - b. Encontre a tarefa “AdVent Technology” e use o modelo analítico de vendas da MicroStrategy. Responda as três questões. Solicite orientações ao seu instrutor.
 - c. Encontre o tutorial sobre modelagem dimensional em freedatawarehouse.com. Prepare um resumo dos cinco pontos mais importantes do tutorial.
2. Pesquise sobre a integração de *data warehouses* e GIS. Comece por mapinfo.com.
3. Faça um teste com os demos de BA da BrightStor, Computer Associates, Temtec, Hyperion e Cognos. Prepare um relatório sobre esses produtos.
4. Considere o problema enfrentado pela cidade de Londres. Desde 17 de fevereiro de 2003, a cidade instituiu uma taxa de entrada no distrito central da cidade para carros e caminhões. Cerca de 1.000 câmeras fotografam digitalmente a placa de cada automóvel que passa. Computadores lêem os números das placas e os combinam com registros em um banco de dados de carros para os quais a taxa foi paga para aquele dia. Se a correspondência não for encontrada, o proprietário do carro recebe uma multa por *email*. As multas variam entre US\$128 e US\$192, dependendo de quando elas são pagas. Verifique os itens relativos a como isso é feito, os erros cometidos e o tamanho do banco de dados envolvido, incluindo aquele com imagens das placas. Também verifique quão bem funciona o sistema através da investigação dos informes jornalísticos. Por fim, relacione o exercício ao OLAP, à BA e OLAP. (Esse exercício foi inspirado em R. Hutton, “London on \$8 a Day!” *Car and Driver*, August 2003, pp. 130–131, caranddriver.com/article.asp?section_id=4&article_id=6859)
5. Realize uma investigação sobre análise de acesso na Web. Compare o Google Analytics ao Web Trend e a outros produtos.
6. Leia “Business Analytics” na *Baseline* de 23 de maio de 2005 em baselinemag.com. Verifique as histórias de sucesso. Por que você acha que a BI é um dos principais projetos de 2005?
7. Acesse businessintelligence.ittoolbox.com. Identifique todos os tipos de *software* de BA. Participe de um grupo de discussão sobre os assuntos abordados neste capítulo. Prepare um relatório sobre suas descobertas.

Tarefas em grupo e interpretação de papéis

1. Acesse **temtec.com** e vá até visão geral Executive Viewer.
 - a. Faça o *tour* guiado. Interaja com cada recurso. Escreva um relatório sobre sua experiência.
 - b. Faça o teste ao vivo e crie cinco consultas e visualizações. Escreva um relatório.
 - c. Teste o *software* com seus próprios dados.
2. A visualização de dados é oferecida por todos os principais fornecedores de BI, e também por outras empresas, como a ILOG (**ilog.com**) e a Crystal Xcelsius (**xcelsius.com**). Cada aluno deve escolher um fornecedor para descobrir os produtos e seus recursos. (Para listas de fornecedores, acesse **tdwi.org** e **dmreview.com**.) Cada grupo deve resumir os produtos e seus recursos.
3. Entreviste administradores em sua faculdade ou executivos em sua empresa a fim de determinar como *data warehousing*, *data mining*, OLAP e ferramentas de BI para visualização poderiam ajudá-los no trabalho. Escreva uma proposta descrevendo suas descobertas. Inclua estimativas de custo e benefícios no seu relatório.
4. Acesse **dmreview.com/resources/demos.cfm**. Verifique a lista de demos e identifique o *software* com recursos analíticos. Cada grupo deve preparar um relatório com pelo menos cinco empresas.
5. Acesse **sas.com** e procure histórias de sucesso relacionadas à BI. Encontre cinco histórias que incluam vídeos SAS e prepare um resumo de cada para apresentação em aula.
6. Prepare um relatório sobre fornecedores de análise da Web. Cada grupo deve apresentar os recursos de duas empresas, como por exemplo, Digital River, WebSideStory, Omniture, ClickTracks e NetIQ.
7. Use o site **google.com** para descobrir aplicações que combinam GIS/GPS. Também procure nos *sites* do fornecedor histórias de sucesso relacionadas a aplicações GIS/GPS. (Para fornecedores de GPS, procure em **biz.yahoo.com** e **google.com**.) Cada grupo deve fazer uma apresentação de cinco aplicações e seus benefícios.
- bloomberg.com**) oferece um serviço personalizado gratuito que mostra o status dos investidores desejados ou listas atuais de ações, incluindo lucros (perdas) e preços (com um atraso de 15 minutos ou até mesmo em tempo real). Como essa informação individualizada é obtida tão rápido? Por que esses dados precisam ser atualizados tão rápido?
2. Encontre casos recentes de aplicações de BA bem-sucedidas. Tente os fornecedores de BI e procure casos ou histórias de sucesso (p.ex., **sap.com**; **businessobjects.com**; **microstrategy.com**). O que você encontrou em comum nas várias histórias de sucesso? Como elas se diferenciam?
3. Acesse *sites* de BI (p.ex., MicroStrategy, Oracle, Hyperion, Microsoft, SAP, SAS, SPSS, Cognos, Temtec, Business Objects) e procure as ferramentas de BA oferecidas. Compare as principais ferramentas e liste seus recursos.
4. Acesse **olapreport.com** e leia sobre a relação do OLAP com as 12 regras de Codd. Prepare um relatório.
5. Acesse **ilog.com/products/optimization** e identifique os produtos de BA (recuperação de *backup* e serviços de mídia, BRMS). Prepare uma lista dos seus recursos.
6. Acesse **fairisaac.com** e encontre produtos para detecção de fraude e análise de risco. Prepare um relatório sobre suas descobertas.
7. Acesse **spss.com** e encontre o *demo* de análise preditiva. Escreva um resumo da utilização e dos benefícios.
8. Acesse **baselinemag.com** e encontre uma lista de mais de 100 produtos de BI. Identifique 10 para BA.
9. Acesse **sas.com** e encontre seus produtos relacionados à BI. Prepare uma lista dos seus recursos.
10. Acesse **microsoft.com/office/dataanalyzer/evaluation/tour**. Faça o *tour* de quatro partes. Resuma os principais recursos do Data Analyzer em um relatório.
11. Acesse **ibm.com** e encontre todos os produtos e serviços de BI da empresa (veja o *demo* de BI). Prepare um relatório sobre suas descobertas.
12. Acesse **navteq.com**. Analise os produtos e as aplicações da empresa. Prepare um relatório sobre cinco aplicações.

Exercícios na Internet

1. A corretagem de ações da America Online e de muitos outros portais (p.ex., **money.cnn.com**;

CASO DE APLICAÇÃO DE FINAL DO CAPÍTULO

Governos estaduais compartilham informações geoespaciais

O GeoStor (geostor.arkansas.gov) é um sistema de GIS público operado pelo estado do Arkansas. Ele inclui mapeamento, gráficos, pesquisa de terras e imóveis e outros dados públicos sobre o Arkansas. É um sistema empresarial que possui seu próprio *data warehouse* (Oracle Applications Server 10g) e um dispositivo especial (Oracle MapViewer) para visualização de dados geoespaciais. O principal uso do sistema é ajudar empresas a tomar decisões sobre onde localizar (ou transferir) instalações no estado. De acordo com o Sistema de Informação Geográfica do Arkansas (relatado por Wiseth, 2004), um interessado do setor industrial com uma lista rigorosa de critérios – distância para transporte, incluindo ferrovias e estradas, e disponibilidade de água, eletricidade e empresa concessionária de gás – estava em processo de avaliação de um terreno no Arkansas como nova localização para seus negócios. Os dados disponíveis no GeoStor foram usados para reunir informação de mão-de-obra, recursos educacionais, recursos de saúde e numerosos outros detalhes, fornecendo um pacote completo de informações dentro do limite de 24 horas exigido pelo interessado. O resultado foi um novo negócio em Osceola, Arkansas – e 500 novos postos de trabalho naquela comunidade.

O GeoStor foi desenvolvido o Arkansas, mas agora está sendo instalado em diversos outros estados nos EUA. Ele dá suporte a muitas outras atividades diárias envolvidas na administração de um estado: tudo desde melhoria na educação até fornecimento de informações de localização fundamentais e atualizadas para resposta a emergências e desastres naturais. O GeoStor está sendo usado para identificar as localizações de todos os recursos educacionais no Arkansas e para mapear essa informação com as métricas de desempenho escolar de modo que as autoridades do Departamento de Educação do Estado possam visualizar como os recursos educacionais estão funcionando hoje e planejar o futuro da educação do Arkansas.

BENEFÍCIOS DO SISTEMA

O GeoStor dá suporte a criação de políticas e tomada de decisão, e também economiza o dinheiro do estado (mais de US\$2 milhões nos primeiros 18 meses). A economia resultou da redução no tempo de busca por documento, de 23,5 horas para 1,5 horas. Uma das principais razões de o tempo ter sido reduzido tão significativamente é porque o trabalho com dados geoespaciais em sistemas tradicionais exige muito tempo e muita mão-de-obra. Por exemplo, o GIS típico é ba-

seado em mapa ou arquivo: usuários fazem *download* de um arquivo inteiro de uma determinada área ou seção de um mapa e, então, trabalham para extrair apenas o que precisam; reciprocamente, os usuários precisam reunir uma profusão de arquivos que abrangem a área de interesse. Por outro lado, o GeoStor possibilita aos usuários pegar apenas a área que precisam. Talvez as questões de qualidade de vida sejam mais importante que economia de custos e eficiência melhorada. Como o sistema fornece um único recurso de veracidade para vários órgãos e dá suporte à interoperabilidade de todas as aplicações clientes que precisam usá-la, a informação no *data warehouse* do GeoStor pode ser mantida totalmente atual – um fato que pode ser traduzido em vidas salvas. Quando há um desastre, é importante que todas as entidades estaduais de infra-estrutura estejam atualizadas; se uma ponte existir, pode significar a diferença entre vida e morte para a equipe de emergência.

Por fim, como descrito anteriormente, o sistema GeoStor fornece vantagem competitiva ao estado do Arkansas em relação aos outros estados. Ele também pode fornecer vantagem competitiva às empresas que usam o sistema para tomar decisões sobre localizações.

Fontes: compilado de K.Wiseth, “The Expanding Role of Location”, *Oracle Magazine*, January/February 2004, and geostor.arkansas.gov (acessado em fevereiro de 2006). Para informações adicionais sobre o tema, acesse opengis.org e otn.oracle.com/products/spatial. Para fazer o *download* do Oracle MapViewer 10g Preview gratuitamente, acesse otn.oracle.com/software/htdocs/devlic.html?/software/products/spatial/htdocs/winsoft.html.

QUESTÕES SOBRE O CASO

1. Por que é necessário um *data warehouse* no sistema GeoStor?
2. Liste os principais benefícios do GeoStor para o Arkansas e para outros usuários do setor privado.
3. Verifique o MapViewer da Oracle e comente sobre seus recursos de visualização de dados.
4. O estado do Arkansas foi o primeiro a ter esse tipo de sistema, por isso tem uma margem competitiva na atração de novos negócios para o estado. Explique por que. Essa vantagem pode ser sustentada quando outros estados começarem a usar um sistema semelhante no futuro?
5. Quaisquer dados de GIS podem ser combinados com dados GPS? Para que usos?

Referências

- "The 2003 CIO Insight Business Intelligence Research Study: Are Your BI Systems Making You Smarter?" (2003, May 23). *CIO Insight*.
- Alexander, S. (2003, February 24). "Web Site Adds Inventory Control and Forecasting." *Computer World*. cognos.com/news/releases/2002/1119_3.html.
- Amato-McCoy, D.M. (2003, May). "Movie Gallery Mines Data to Monitor Associate Activities." *Stores*.
- Atre, S. (2003, June 30). "The Top 10 Critical Challenges for BI Success." *Computer World*, special advertising supplement.
- Baer, T. (2002, April). "Analyzing Data at Real Time." *Application Development Trends*.
- Barrett, L., and D.F. Carr. (2005, September). "Proud Sponsor of the American Dream." *Baseline*.
- Bolen, A. (2003, Quarter 2). *SAS Cooks Up Success at Williams-Sonoma*, sas.com/success/williamssonoma.html.
- Bonde, A., and M. Kuckuk. (2004, April). "Real-World Business Intelligence: The Implementation Perspective", *DM Review*.
- Brath, R., and M. Peters. (2006, January). "Visualization Spreadsheets." *DM Direct*.
- Callaghan, D. (2005, February 7). "BI Vendors Stress Need for Integration." *eWEEK*.
- Callaghan, D. (2003a, May 26). "SPSS, SAS Take Predictive Paths." *eWEEK*.
- Callaghan, D. (2003b, June 2). "IBM Builds Bridge for DB2." *eWEEK*.
- Codd, E.F., S.B. Codd, and C.T. Salley. (1993, July). "Beyond Decision Support." *Computerworld*.
- Coffee, P. (2003, June 23). "'Active' Warehousing." *eWEEK*.
- Cognos.com, (2002, November 19). "Allied Building Products Corporation Monitors Corporate Performance with Cognos Finance." cognos.com/news/releases/2002/1119_3.html (acessado em outubro de 2006).
- Comcowich, W.J. (2002, October 22). *Integrated Internet Monitoring Solutions for CI*, imakenews.com/scip2/e_article000101312.cfm (acessado em fevereiro de 2006).
- Culpepper, R.B. (2002, October). "Quick-Take Reviews: CommunityViz 1.3." *GeoWorld*.
- Davenport, T.H., and J.G. Harris. (2005, Summer). "Automated Decision Making Comes of Age." *MIT Sloan Management Review*.
- Devlin, B. (2003, May 14). "Solving the Data Warehouse Puzzle." *DB2 Magazine*, available at datawarehouse.ittoolbox.com/documents/industry-articles/solving-the-data-warehouse-puzzle-1583 (acessado em março de 2006).
- Dorrington, P. (2003, Quarter 2). *Innovative, Industrious and Nefarious Fraudsters!* sas.com/news/sascom/2003q2/feature_fraudsters.html (acessado em fevereiro de 2007).
- Eckersen, W. (2003). *Smart Companies in the 21st Century*. Seattle, WA: The Data Warehousing Institute.
- Edelstein, H.A. (2001, March 12). "Pan for Gold in the Clickstream." *InformationWeek*.
- Fayyad, U., G. Grinstein, and A. Wierse (eds). (2002). *Information Visualization in Data Mining and Knowledge Discovery*. San Francisco: Morgan Kaufman.
- Fleischmann, M., J.M. Hall, and D.F. Pyke. (2004, Winter). "Smart Pricing." *MIT Sloan Management Review*.
- Fogarty, K. (2004, December 1). "Primer: Predictive Analytics." *Baseline*.
- Follett, J.H. (2003, June 23). "HP Gives Partners an Open View of Adaptive Enterprises." *CRN*.
- Franklin, D. (2002, November). "Any Way You Slice It." *Credit Union Management*.
- Gage, D. (2005, December 13). "Forget Hackers: Watch Out for Competitors", *Baseline*, baselinemag.com/article2/0,1397,1901400,00.asp (acessado em março de 2006).
- Gonzales, M.L. (2003, February 1). "The New GIS Landscape." *Intelligent Enterprise*.
- Gray, P., and H.J. Watson. (1998). *Decision Support in the Data Warehouse*. Upper Saddle River, NJ: Prentice Hall.
- Hallett, P. (2001, June). "Web-Based Visualization." *DM Review*.
- Hamilton, J. M. (1996, March 15). "A Mappable Feast." *CIO Magazine*.
- Havenstein, H. (2003a, April 28). "Real-Time Smarts." *InfoWorld*.
- Havenstein, H. (2003b, May 26). "BAM Bolsters Data Visibility." *InfoWorld*.
- Ho, G.T.S., H.C.W. Lau, W.H. Ip, and A. Ning. (2004, July). "An Intelligent Information Infrastructure to Support the Streamlining of Integrated Logistics Workflow." *Expert Systems*.
- Hoffer, J.A., M.B. Prescott, and F.R. McFadden. (2005). *Modern Database Management*, 7th ed. Upper Saddle River, NJ: Prentice Hall.
- "Houston Buses Due for 'Intellectual Overhaul.'" (2003, June). *OR/MS Today*, p. 19, lionhrtpub.com/orms/news/realworld/rw603buses.html (não está mais disponível online).
- Infrastructure and Management Solutions for the Adaptive Enterprise*. (2003). hp.com/products1/promos/adaptive_enterprise/pdfs/vision_for_ae.pdf (acessado em fevereiro de 2006).
- Langnau, L. (2003, November). *Business Intelligence and Ethics: Can They Work Together? Controls & Systems Editorial—Industry Overview*, mhmonline.com/nID/1180/iID/20865/MHM/viewStory.asp (acessado em agosto de 2006).
- Lindquist, C. (2003, May 15). "Real Timing." *CIO*.
- McCarthy, J., and M. McCarthy. (2002). *Software for Your Head: Core Protocols for Creating and Maintaining Shared Vision*. Boston: Addison-Wesley.
- McKinley, E. (2004, November). "We're Not Asking, We Know You Want Fries with That Predictive Technology." *Stores*.
- McKnight, W. (2005, February). "Building Business Intelligence: Will Business Intelligence Replace the Busi-

- ness Analyst?” *DM Review*, dmreview.com/article_sub.cfm?articleId=1018117 (acessado em março de 2006).
- Meier, M., W. Sigzig, and P. Mertens. (2005). *Enterprise Management with SAP SEM/Business Analytics*, 2nd ed. Berlin: Springer.
- MicroStrategy. (2006). *The 5 Styles of Business Intelligence*, microstrategy.com/Solutions/5Styles/enterprise_reporting.asp (acessado em fevereiro de 2007).
- MicroStrategy. (2005). *Applications of Industrial-Strength Business Intelligence*, microstrategy.com/Solutions/AppsBook.asp (acessado em agosto de 2006).
- Morris, H., S. Graham, P. Andersen, K. Moser, R. Blumstein, D. Vesset, N. Martinez, and M. Carr. (2002, October). “The Financial Impact of Business Analytics.” *IDC*.
- Moss, L. T., and S. Atre. (2003). *Business Intelligence Roadmap*. Indianapolis: Addison-Wesley.
- Pallatto, J. (2002, February). “Business Tools Get Smart.” *InternetWorld*.
- Patel, J. (2005, May). “Seven Simple Rules for Successful Real-Time Business Intelligence.” *DM Review*.
- Pawar, S. P., and R. Sharda. (1997). “Obtaining Business Intelligence on the Internet.” *Long Range Planning*, Vol. 30, No. 1.
- Perry, J. J., and G. V. Post. (2007). *Introduction to Oracle 10G*. Upper Saddle River, NJ: Prentice Hall.
- Peterson, E. (2005, August). *Web Analytics Demystified*. Sebastopol, CA: O’Reilly Media.
- Peterson, T. (2003, April 21). “Getting Real About Real Time.” *Computerworld*.
- Raden, N. (2003a, June 17). “Real Time: Get Real, Part I.” *Intelligent Enterprise*.
- Raden, N. (2003b, June 30). “Real Time: Get Real, Part II.” *Intelligent Enterprise*.
- Rapoza, J. (2004, June 7). “Site Analysis Gets Expert Treatment.” *eWEEK*.
- Ruber, P. (2003, June). “Analytics Improve Merchandising.” *InternetWorld*.
- Saarevirta, G. (2004, Winter). “The Untapped Value of GIS”, *Business Intelligence Journal*.
- Schlegel, K. (2003, June). “Web Analytics Essentials.” *InternetWorld*.
- Schwartz, M. (2005, Fall). “Louisiana Bottles Food Stamp Fraud.” *Business Intelligence Journal*.
- Swoyer, S. (2005, Summer). “BI Case Study: Summit Partners.” *Business Intelligence Journal*.
- Thompson, O. (2004, October 9). *Business Intelligence Success, Lessons Learned*. bettermanagement.com/library/library.aspx?libraryid=1675 (acessado em fevereiro de 2007).
- Thompson, O., and P. J. Jakovljevic. (2005, June). *Business Intelligence Status Report* (relatório compreensivo de sete partes, incluindo capacidades em tempo real e on-demand), TechnologyEvaluation.com.
- Turban, E., et al. (2006). *E-Commerce: A Managerial Perspective*. Upper Saddle River, NJ: Prentice Hall.
- Ursery, S. (2004, February). “GIS More Prevalent in Big Cities”, *The American City and County*.
- Werner, V., and C. Abramson. (2001, Summer). “Managing Clickstream Data.” *Journal of Data Warehousing*, Vol. 6, No. 3.
- White, C. (2004, September). “Now Is the Right Time for Real-Time BI.” *DM Review*.
- Wiseth, K. (2004, January 14). “The Expanding Role of Location.” *Oracle Magazine*. oracle.com/technology/oramag/oracle/04-jan/o14geostore_feature.html (acessado em fevereiro de 2006).
- Zaman, M. (2005, November 8). “Predictive Analytics; the Future of Business Intelligence.” *Technology Evaluation Centers*.



Data, Text e Web Mining

Objetivos de aprendizado

- ◆ Descrever *data mining* e listar seus objetivos e benefícios
- ◆ Entender diferentes finalidades e aplicações de *data mining*
- ◆ Entender diferentes métodos de *data mining*, em especial modelos de agrupamento e árvore de decisão
- ◆ Desenvolver habilidade no uso de alguns *softwares* de *data mining*
- ◆ Aprender o processo dos projetos de *data mining*
- ◆ Entender armadilhas e mitos do *data mining*
- ◆ Descrever *text mining* e seus objetivos e benefícios
- ◆ Avaliar o uso de *text mining* em aplicações comerciais
- ◆ Descrever *Web mining* e seus objetivos e benefícios

D*ata mining* refere-se ao desenvolvimento do *business intelligence* a partir dos dados coletados, organizados e processados por uma empresa. Técnicas de *data mining* estão sendo usadas por empresas para obter um melhor entendimento dos seus clientes e de suas próprias operações. Neste capítulo, estudaremos algumas das principais técnicas de *data mining*, como por exemplo, árvores de decisão e análise de *clusters*, suas aplicações e os processos de *data mining*.

- 4.1 Vinheta de abertura: Highmark Inc. emprega *data mining* para gerenciar as despesas com seguros
- 4.2 Conceitos e aplicações de *data mining*
- 4.3 Técnicas e ferramentas de *data mining*
- 4.4 Processos do projeto de *data mining*
- 4.5 *Text mining*
- 4.6 *Web mining*

4.1

VINHETA DE ABERTURA: HIGHMARK INC. EMPREGA DATA MINING PARA GERENCIAR AS DESPESAS COM SEGUROS

A Highmark Inc., estabelecida em Pittsburgh, tem uma longa tradição no fornecimento de acesso a serviços de saúde de qualidade com preços acessíveis aos seus membros e a sua comunidade.

A empresa foi criada em 1996 pela fusão de duas licenciadas da Blue Cross e da Blue Shield Association: a Blue Shield Pensilvânia (agora Highmark Blue Shield) e um plano da Blue Cross na região oeste da Pensilvânia (agora Highmark Blue Cross Blue Shield). Atualmente, a Highmark é uma das maiores seguradoras de saúde nos Estados Unidos.

Dados em empresas de assistência médica

A quantidade de dados que circulam nas empresas de assistência médica como a Highmark é enorme. Esses dados, muitas vezes considerados como ocupantes de espaço de armazenamento e pensados como uma ameaça com a qual se deve lidar, oferecem novos usos interessantes. Empresas de *data mining* fornecem ferramentas úteis para análise de dados do paciente e descoberta de mistérios que podem levar a uma melhor assistência médica com baixos custos – uma tarefa que a maioria das empresas de assistência médica está tentando realizar.

A cada dia, essas empresas recebem milhões de dados sobre seus clientes, e cada pedaço de informação atualiza a história de caso de cada membro. As empresas tornaram-se conscientes da utilidade dos dados a sua disposição e usam ferramentas para geração de relatórios de dados e outros *softwares* analíticos a fim de extrair agrupamentos de pacientes que são mais dispendiosos no tratamento do que a maioria. Esforços anteriores no uso de tecnologia computacional para extrair informações de pacientes eram limitados ao estabelecimento de uma ligação entre duas doenças. O *software* poderia, por exemplo, percorrer dados e relatar que diabéticos ou pessoas com doenças coronárias têm o tratamento mais caro. Entretanto, *software* e ferramentas para geração de relatórios anteriores não eram eficientes para descobrir por que esses pacientes adoeciam ou por que alguns pacientes eram mais afetados por determinadas doenças do que outros. Ferramentas de *data mining* podem resolver alguns dos problemas através da análise das informações produzidas e da criação de relações e correlações concisas.

Empregar *data mining* é caro e complexo. Também depende da eficiência dos dados processados por uma empresa. Primeiramente, as empresas precisam investir em *data warehousing* para gerenciar dados volumosos. O *data warehousing* precisa de monitoramento e manutenção contínuos, o que acarreta em mais investimento. A maioria das empresas não está integrada em um nível considerável; elas ainda usam vários servidores e *warehouses*. A presença de múltiplos *warehouses* dificulta a integração da informação porque, com frequência, o formato dos dados em sistemas variados é diferente. As empresas também enfatizam a importância da consistência na tentativa de reunir informações; porém, elas estão concentradas na coleta de dados relevantes para suas operações. Conflitos internos de gerenciamento também podem impedir o uso de aplicações de *data mining*.

Empresas de assistência médica são inundadas com dados e algumas delas não querem contribuir com a complexidade ao adicionar aplicações de *data mining*. Talvez elas queiram percorrer dados por diversas razões, mas são incapazes de decidir por que e como analisar seus dados. De qualquer modo, as coisas estão ficando mais claras para os pacientes assim como para as empresas, pois as regras dos seguros de saúde estão abrindo caminho para análises estruturais e de dados eficientes.

A necessidade de *data mining*

As pressões de mercado estão fazendo com que as empresas de assistência médica levem o *data mining* a sério. Clientes e concorrentes estão focados, o que resulta em produtos mais personalizados. Essa personalização nos leva ao ponto de origem do por que e onde acontecem os custos médicos. Muitas empresas começaram a usar *softwares* preditivos para prever quem está propenso a adoecer e para quem o tratamento será caro no futuro. Uma olhada no futuro permitiu às empresas filtrarem seus pacientes caros e reduzir seus custos de Medicare através do uso de medidas preventivas. Outra aplicação importante dos estudos preditivos é o controle de bônus. Um grupo empregador que tem um grande número de funcionários pertencentes a uma faixa de alto custo veria suas taxas crescerem.

A modelagem preditiva pode ser capaz de prever quais pacientes estão propensos a adoecer no futuro. Uma aplicação de modelagem preditiva pode classificar um paciente diabético como um alto risco de custos médicos elevados. Entretanto, o *data mining* usado na Highmark traçar a uma relação entre um paciente diabético e outros parâmetros. Um paciente com uma condição cardíaca específica pode, por exemplo, ter um alto risco de ter diabetes; essa relação é traçada porque a medicação cardíaca pode levar o paciente a desenvolver essa doença posteriormente. Executivos da Highmark atestaram esse fato dizendo que não teriam monitorado pacientes quanto ao medicamento cardíaco e nem teriam traçado a relação entre a medicação cardíaca e o diabetes. A pesquisa médica tem sido bem-sucedida em sistematizar muitas das complexidades associadas às condições do paciente. O *data mining* estabeleceu as bases para os programas de intervenção.

Fontes: resumido de G. Gillespie, “Data Mining: Solving Care, Cost Capers”, *Health Data Management*, November 2004, findarticles.com/p/articles/mi_km2925/is_200411/ai_n8622737 (acessado em abril de 2006); and *Highmark Enhances Patient Care, Keeps Medical Costs Down with SAS*, sas.com/success/highmark.html (acessado em abril de 2006).

Questões sobre a vinheta de abertura

1. Por que empresas como a Highmark usam aplicações de *data mining*?
2. Por que, no início, as empresas de assistência médica estavam hesitantes quanto ao uso de aplicações de *data mining*?
3. Quais são as possíveis ameaças que poderiam surgir com as aplicações de *data mining*?
4. Quais complexidades surgem quando *data mining* é usado em empresas de assistência médica?
5. Suponha que você é funcionário e que sua empresa de assistência médica aumenta sua taxa com base nos resultados de *software* de *data mining* e modelagem preditiva. Você aceitaria os prognósticos feitos pelo *software* da empresa?

O que podemos aprender com esta vinheta?

A área da saúde é a primeira candidata para *data mining*. A quantidade de dados gerados pela área da saúde é volumosa, mas apenas uma pequena fração dela está sendo explorada a fim de desenvolver profundo entendimento das inter-relações em questões de saúde. Neste capítulo, veremos outras aplicações de *data mining* na área da saúde, em conjunto com aplicações em muitas outras áreas. Essas aplicações criam questões éticas e legais que devem ser abordadas pelas empresas.

4.2

CONCEITOS E APLICAÇÕES DE DATA MINING

Em um artigo recente na *Harvard Business Review*, Davenport (2006) afirmou que a mais nova arma estratégica para as empresas é a tomada de decisão analítica. Ele forneceu exemplos de empresas, como a Amazon.com, Capital One, Marriot International entre outras, que têm usado análise para melhor entender seus clientes e otimizar suas cadeias de fornecimento ampliadas para maximizar seu retorno de investimento, ao mesmo tempo em que fornece o melhor serviço ao cliente. Esse nível de sucesso é altamente dependente de quão bem uma empresa entende seus clientes, fornecedores, sua cadeia de fornecimento e assim por diante. Um grande componente desse entendimento vem da análise dos dados coletados por uma empresa. O custo de armazená-los e processá-los diminuiu drasticamente nos últimos anos e, como resultado, a quantidade de dados armazenados em formatos eletrônicos cresceu em uma velocidade explosiva. Com a criação de grandes bancos de dados, surgiu a possibilidade de analisar os dados armazenados neles. Originalmente, o termo *data mining* era usado para descrever o processo no qual padrões anteriormente desconhecidos eram identificados nos dados. Desde então, essa definição tem sido estendida para além desses limites, de modo a incluir a maioria dos modelos de análise de dados. Como conse-

qüência, o rótulo *data mining* muitas vezes é usado para adicionar valor de vendas a quase todo tipo de ferramentas de análise de dados.

Embora o termo *data mining* seja relativamente novo, as idéias por trás dele não são. Muitas das técnicas usadas em *data mining* têm suas bases na análise estatística tradicional e no trabalho de inteligência artificial dos anos 80. Então, por que repentinamente esse processo ganhou a atenção do mundo dos negócios? Cavoukian (1998) cita seis fatores de um relatório da IBM após esse repentino crescimento de popularidade:

1. Reconhecimento geral do valor inexplorado em grandes bancos de dados
2. Consolidação dos registros em banco de dados, levando a uma única visão do cliente
3. Consolidação dos bancos de dados, incluindo o conceito de repositório de informações
4. Redução no custo de armazenamento e processamento de dados, sustentando a capacidade de coletar e conservar dados
5. Concorrência intensa pela atenção do cliente em um mercado cada vez mais saturado
6. Movimento em direção a desmassificação (conversão dos recursos de informação em formas não-físicas) das práticas de negócio

Áreas de estudo como astronomia e física nuclear armazenam grandes quantidades de dados em bases de rotina. Os dados na Internet estão crescendo em volume e complexidade. Grandes volumes de dados genômicos estão sendo reunidos em todo o mundo. Pesquisadores médicos e farmacêuticos usam aplicações de *data mining* para identificar terapias bem-sucedidas para doenças e descobrir medicamentos novos e melhorados.

No lado comercial, talvez o uso mais comum de *data mining* seja nos setores financeiro, varejista e da saúde. O *data mining* é usado para reduzir comportamento fraudulento, particularmente em reivindicações de seguro e uso de cartão de crédito (Chan et al., 1999); para identificar padrões de compra do cliente (Hoffman, 1999); para recuperar clientes rentáveis (Hoffman, 1998); para identificar regras de negócio a partir de dados históricos; e para auxiliar na análise de cesta de supermercado. O *data mining* já é amplamente usado para melhor visar clientes, e, com o desenvolvimento do comércio eletrônico, a tendência é que isso se torne mais importante com o passar do tempo.

Definições, características e benefícios

Data mining é o termo usado para descrever a descoberta de informações em bancos de dados. O *data mining* é um processo que usa técnicas estatísticas, matemáticas, de inteligência artificial e de aprendizagem automática para extrair e identificar informações úteis e conhecimento subsequente de bancos de dados. Antigamente, o termo era usado para descrever o processo no qual padrões desconhecidos eram identificados nos dados. Entretanto, ao longo do tempo, a definição original foi modificada para incluir a maioria dos tipos de análise (automática) de dados. *Data mining* é o processo de descoberta de padrões matemáticos em grandes conjuntos de dados, geralmente. Esses padrões podem ser regras, semelhanças, correlações, tendências ou modelos preditivos (veja Nemati e Barko, 2001).

O *data mining* tem sua base na interface da ciência da computação e da estatística, usando avanços nas duas disciplinas para progredir na extração de informações de grandes bancos de dados. É um campo emergente que tem atraído muita atenção em muito pouco tempo. Glymour et al. (1997) discutiram temas e lições estatísticas diretamente relevantes ao *data mining*, assim como oportunidades de sinergia entre as comunidades computacional e estatística para avanços futuros na análise de dados.

O *data mining* inclui tarefas como extração de conhecimento, arqueologia de dados, exploração de dados, processamento de padrões de dados, limpeza de dados e colheita de informação.

Todas essas atividades são administradas automaticamente e permitem uma descoberta rápida até mesmo por não-programadores (veja Caso de aplicação 4.1). A seguir são apresentadas as principais características e objetivos do *data mining*:

- Muitas vezes, os dados estão escondidos em bancos de dados muito grandes que, às vezes, contêm dados de diversos anos. Em muitos casos, os dados estão limpos e consolidados em um *data warehouse*.
- O ambiente de *data mining* geralmente é uma arquitetura cliente/servidor ou uma arquitetura baseada na Web.
- Novas ferramentas sofisticadas, incluindo ferramentas de visualização avançada, ajudam a remover informações escondidas em arquivos corporativos ou arquivadas em registros públicos. Descobri-las envolve mexer e sincronizar os dados para a obtenção dos resultados certos. *Data miners* modernos também estão explorando a utilidade dos dados temporários (ou seja, texto não estruturado armazenado em lugares como bancos de dados Lotus Notes, arquivos de texto na Internet ou nas *intranets* de toda a empresa).
- Muitas vezes, o *miner* é um usuário final, capacitado por *datadrills* e outras ferramentas poderosas de consulta para fazer perguntas *ad hoc* e obter respostas rapidamente, requerendo pouca ou nenhuma habilidade de programação.
- Tirar a sorte grande muitas vezes envolve descobrir um resultado inesperado e exige que os usuários finais pensem de forma criativa.
- Ferramentas de *data mining* são facilmente combinadas com planilhas e outras ferramentas para desenvolvimento de *software*. Dessa forma, os dados extraídos podem ser analisados e processados de maneira rápida e fácil.
- Devido às grandes quantidades de dados e iniciativas sólidas de pesquisa, às vezes é necessário usar processamento paralelo para *data mining*.

CASO DE APLICAÇÃO 4.1

Dados ajudam a prever as necessidades do cliente

Os videntes do First Health Group Corp. (FHG Corp.) estão ocupados descobrindo quais questões e preocupações seus clientes podem ter sobre seus planos de saúde. Os executivos podem prever o que os membros do plano de saúde precisam através do uso de aplicações de *data mining* que investigam seus bancos de dados. O FHG Corp. criou regras baseadas em um mecanismo de *data mining* para entender e documentar ligações telefônicas feitas pelos seus clientes a fim de prever que tipo de questão ou preocupação alguém pode ter.

O FHG Corp. emprega tecnologia de *data mining* para coletar informações dos seus sistemas centrais de informação e as apresenta ao pessoal do atendimento por meio de uma aplicação de gerenciamento de relacionamento com o cliente (CRM) exclusiva da empresa. A aplicação de *data mining* cria uma lista de ligações a serem feitas, a cada manhã, pelo funcionário do atendimento ao cliente.

As regras são baseadas nos tipos de questões que os membros fazem sobre benefícios, ações e outros serviços. Então, essas regras são processadas em comparação com a informação mais recente nos bancos de dados. O mecanismo de regra percorre os dados em vários critérios e cria arquivos. Em seguida, a equipe de atendimento ao cliente usa os arquivos para contatar os membros quanto às várias questões relacionadas a cuidados médicos e futuras necessidades de saúde.

Alguns dos benefícios desse sistema incluem redução no tempo de trâmite das ações e um decréscimo de 65 a 70% no número de ações pendentes.

Fontes: B. Briggs, "Data Helps Foretell Customer Needs", *Health Data Management*, Vol. 12, No. 2, February 2004, and *Health Data Management*, **Keepmedia.com**, February 2, 2004.

Uma empresa que potencializa efetivamente ferramentas e tecnologias de *data mining* pode adquirir e manter uma vantagem competitiva estratégica. O *data mining* oferece às empresas um ambiente indispensável de melhoria da decisão para aproveitar novas oportunidades através da transformação dos dados em arma estratégica. Veja Nemati e Barko (2001).

Como funciona o *data mining*

Segundo Dunham (2003), um *data mining* inteligente descobre informações em *data warehouses* onde consultas e relatórios não conseguem revelar efetivamente. Ferramentas de *data mining* encontram padrões em dados e podem até deduzir regras a partir deles. Três métodos são usados para identificar padrões em dados (veja Nemati e Barko, 2001):

- Modelos simples (p.ex., consultas baseadas em SQL, processamento analítico *online* [OLAP], raciocínio humano)
- Modelos intermediários (p.ex., regressão, árvores de decisão, agrupamento)
- Modelos complexos (p.ex., redes neurais, outra indução de regras)

Esses padrões e regras podem ser usados para direcionar a tomada de decisão e prever as consequências das decisões. O *data mining* pode acelerar a análise ao concentrar a atenção nas variáveis mais importantes. A queda drástica na relação custo-desempenho de sistemas computacionais permitiu às empresas iniciar a aplicação de algoritmos complexos de técnicas de *data mining*. Tradicionalmente, os algoritmos de *data mining* podem ser divididos em quatro categorias amplas: classificação, agrupamento, associação e descoberta de sequência. Outras ferramentas de análise de dados, como regressão e análise de séries temporais, também encontram o caminho dentro da prática, assim como a visualização. Neste livro, temos uma visão geral de todas as categorias, mas aprenderemos com mais detalhes sobre classificação e agrupamento. O Capítulo 6 (*online*) abrange redes neurais, um paradigma diferente de modelagem para uso em *data mining*.

Classificação

Classificação, ou indução supervisionada, talvez seja a mais comum de todas as atividades de *data mining*. O objetivo da classificação é analisar os dados históricos armazenados em um banco de dados e gerar automaticamente um modelo que possa prever comportamento futuro. Esse modelo induzido consiste de generalizações sobre os registros de um conjunto de dados de treinamento, que ajuda a distinguir classes redefinidas. A expectativa é que o modelo possa ser usado para prever as classes de outros registros não-classificados.

Ferramentas comuns usadas para classificação são redes neurais, árvores de decisão e regras “se-então-senão” que não precisam ter uma estrutura em árvore. Redes neurais (veja Capítulo *online* 6) envolvem o desenvolvimento de estruturas matemáticas que têm a capacidade de aprender. Elas tendem a ser mais eficientes onde o número de variáveis envolvido é muito grande e as relações entre elas é complexa e imprecisa. Uma rede neural pode ser facilmente implementada em um ambiente paralelo, com cada nó da rede fazendo seus cálculos em um processador diferente. Redes neurais têm vantagens e desvantagens. Por exemplo, normalmente é muito difícil fornecer uma boa justificativa para as previsões feitas por uma rede neural. Além disso, redes neurais tendem a precisar de muito treinamento. Infelizmente, o tempo necessário para treinamento tende a aumentar com o aumento do volume de dados e as redes neurais, em geral, não podem ser treinadas em bancos de dados muito grandes. Esses e outros fatores têm limitado a aceitabilidade dessas redes.

Árvores de decisão classificam dados em um número finito de classes, com base nos valores das variáveis de entrada. Árvores de decisão são compostas basicamente de uma hierarquia de declarações se-então e, portanto, são significativamente mais rápidas do que as redes neurais. Elas são mais adequadas para dados categorizados e intervalares, pois incorporar variáveis contínuas

em uma estrutura de árvore de decisão pode ser difícil. Uma ferramenta relacionada à classificação é a indução de regras; ao contrário de uma árvore de decisão, as declarações “se-então” usadas não precisam ser hierárquicas.

Agrupamento

O **agrupamento** divide um banco de dados em segmentos cujos membros compartilham qualidades semelhantes. Algumas das idéias usadas para classificação, como redes neurais, referem-se em parte a situações que envolvem agrupamento. Porém, diferente da classificação, os *clusters* são desconhecidos no agrupamento quando o algoritmo começa. Portanto, antes dos resultados das técnicas de agrupamento serem usadas, talvez seja necessário a interpretação por parte de um especialista, com possível modificação, dos *clusters* sugeridos. Após os clusters aceitáveis serem identificados, eles podem ser usados para classificar novos dados. Como era de se esperar, as técnicas de agrupamento incluem otimização; a meta é criar grupos para que os membros dentro de cada grupo tenham semelhança máxima e os membros fora dos grupos tenham semelhança mínima.

Associação

Associações estabelecem relações entre itens que ocorrem juntos em um determinado registro. A coleta de dados tem sido drasticamente simplificada como resultado de scanners, e a determinação de associações entre itens que são vendidos juntos pode ser um benefício substancial para os varejistas. Muitas vezes, é chamada de *análise de cesta de supermercado* porque uma das aplicações primárias dessa técnica é a análise das operações de venda.

Descoberta de sequência

Descoberta de sequência é a identificação de associações ao longo do tempo. Quando informações adequadas estão disponíveis (p.ex., a identidade de um cliente em uma loja), uma análise temporal pode ser realizada para identificar o comportamento ao longo do tempo. Algumas técnicas de descoberta de sequência rastreiam o tempo decorrido entre eventos associados e a frequência de sequência sucedida. Isso fornece uma quantidade considerável de informação que pode ser usada para aumentar as vendas ou detectar fraudes.

Visualização

Os *insights* a serem obtidos com a visualização de dados não podem ser supervalorizados. Isso se confirma para a maioria das técnicas de análise de dados, mas é especialmente verdadeiro em *data mining*. Dado o volume absoluto de dados nos bancos de dados em consideração, em geral a visualização é um empreendimento difícil. Entretanto, pode ser usada em conjunto com o *data mining* para obter um entendimento mais evidente de muitas relações subjacentes.

Regressão

Regressão é uma técnica estatística conhecida usada para mapear dados para um valor de previsão. Técnicas de regressão linear e não-linear são usadas. Regressão é uma forma de estimativa. Muitas vezes envolve a identificação de métricas e a avaliação de um item (p.ex., um cliente) junto às métricas através da especificação de pontuações. Previsões de venda também podem ser realizadas.

Previsão

Previsão é outra forma de estimativa. Ela avalia valores futuros com base em padrões dentro de amplos conjuntos de dados (p.ex., previsão de demanda). Existe um esforço para usar métodos estatísticos de séries temporais a fim de prever vendas futuras.

Tabela 4.1 Funções de *data mining*, algoritmos e exemplos de aplicação

<i>Funções de data mining</i>	<i>Algoritmo</i>	<i>Exemplos de aplicação</i>
Associação	Estatística, teoria dos conjuntos	Análise de cesta de supermercado
Classificação	Árvores de decisão, redes neurais, controle, avaliação de risco, regras	Marketing direcionado, qualidade
Agrupamento	Redes neurais, estatística, otimização, análise discriminante	Segmentação de mercado
Descoberta de seqüência	Estatística, teoria dos conjuntos	Análise de cesta de supermercado ao longo do tempo, análise do ciclo de vida do cliente
Modelagem	Regressão linear e não-linear, ajuste da curva, redes neurais	Previsão de vendas, nível de interesse, predição, controle de estoque
Drill-down e visão agregada dos dados	Visualização, usando muitas abordagens diferentes	Quase todas as aplicações anteriores

Fonte: adaptado de J.P. Bigus, *Data Mining with Neural Networks*, McGraw-Hill, New York, 1996.

Resumo das funções de *data mining*

Na Tabela 4.1, mostramos as funções de *data mining* descritas até agora, junto com algoritmos representativos e exemplos de aplicação. Veja também Groth (1998).

Muitas vezes, empresas usam seus sistemas de *data mining* para realizar segmentação de mercado com análise de *cluster*. Essa análise é um meio de identificar classes para que os itens em um *cluster* tenham mais em comum entre si do que com itens em outros *clusters*. Pode ser usada na segmentação de clientes e direcionamento adequado do *marketing* de produtos para os segmentos na hora certa, da forma certa e com o preço certo. Tillet (2000) descreveu como outro banco explora efetivamente os dados do cliente através da análise de *cluster*, usando ferramentas baseadas na Web. O Caso de aplicação 4.2 descreve como a análise de *cluster* foi combinada com outras técnicas de *data mining* para verificar as causas de acidentes.

CASO DE APLICAÇÃO 4.2

Acidentes com veículos automotores e as distrações do motorista

A distração do motorista está na posição central das preocupações com segurança na estrada. Um estudo publicado em 1996 pela NHTSA (National Highway Traffic Safety Administration) concluiu que aproximadamente 25 a 30% dos ferimentos causados por acidentes de carro eram devido à distração do motorista. Em 1999, segundo o FARS (Fatality Analysis Reporting System) desenvolvido pelo NCSA (National Centre for Statistics and Analysis), 11% dos acidentes fatais (ou seja, 4.462 mortes) eram devido à falta de atenção do motorista.

Um estudo foi realizado para extrair os padrões dos fatores de distração e acidentes de trânsito. O *data mining* foi a principal ferramenta usada para

traçar a correlação dos dados a partir das informações sobre acidentes fornecidas pelo FARS. Três técnicas de *data mining* – redes de Kohonen, árvores de decisão e redes neurais para descobrir diferentes combinações de fatores de distração que explicassem os altos índices de acidentes – foram usadas na pesquisa. As redes de Kohonen detectaram *clusters* e revelaram padrões de variáveis de entrada na coleta de dados. As árvores de decisão exploraram e classificaram a consequência de cada incidente em eventos sucessivos; as árvores de decisão sugeriram a relação entre motoristas desatentos e condições físicas/mentais. As técnicas de *data mining* foram aplicadas aos conjuntos de dados a fim de correlacionar falta

de atenção e outros fatores relacionados ao motorista em acidentes de trânsito. Por fim, um modelo de rede neural foi treinado e testado para observar a eficácia do modelo. O Clementine, da SPSS, foi usado para extrair os dados derivados do banco de dados do FARS em três modelos.

O sistema identificou 1.255 motoristas que estiveram envolvidos em acidentes no qual a falta de atenção foi um dos fatores determinantes que levou ao acidente. Colisões frontais e traseiras e outros

tipos de colisão, entre outras diversas variáveis de saída, foram previstas com 78 e 77% de precisão.

Fontes: W. S. Tseng, H. Nguyen, J. Liebowitz, and W. Agresti, "Distractions and Motor Vehicle Accidents: Data Mining Application on Fatality Analysis Reporting System (FARS) Data Files", *Industrial Management & Data Systems*, Vol. 105, No. 9, January 2005, pp. 1188–1205; and J. Liebowitz, "New Trends in Intelligent Systems", Apresentação realizada na Universidade de Granada, docto-si.ugr.es/seminario/presentaciones/jay.ppt (acessado em agosto de 2006).

O *data mining* pode ser tanto *baseado em hipótese* quanto *baseado em descoberta*. O ***data mining* baseado em hipótese** começa com uma proposição pelo usuário, que então busca validar a veracidade da proposição. Um gerente de *marketing* pode começar, por exemplo, com a proposição "As vendas de DVD *player* estão relacionadas às vendas de aparelhos de televisão?". O Caso de aplicação 4.2 ilustra como as técnicas de *data mining* abordadas neste capítulo e no Capítulo 6 podem ser usadas para verificar conjecturas. O ***data mining* baseado em descoberta** encontra padrões, associações e relações entre os dados. Pode revelar fatos que uma empresa não conhecia ou não observou anteriormente. O Caso de aplicação 4.3 descreve um processo para conversão desses dados em informações acionáveis.

CASO DE APLICAÇÃO 4.3

***Data mining* para identificar o comportamento do cliente**

Entender o comportamento do cliente é importante no ajuste das estratégias de negócios, no aumento dos lucros e na identificação de novas oportunidades. Muitas empresas têm uma volumosa quantidade e impressionante variedade de recursos de dados e informações que prometem revelar muita mais sobre o comportamento do cliente do que se pensava que era possível antigamente. Muitas organizações atingiram uma situação de dados valiosos e utilização ineficiente. Para a maioria dos ambientes de varejo, três fontes de dados do cliente são mais importantes para as tentativas de *data mining* visadas ao melhor entendimento do comportamento:

- Dados demográficos
- Dados de operação
- Dados de interação *online*

A análise de acesso na Web pode ser usada para identificar quem comprou e quem não comprou um produto, por que e quando.

O processo de *data mining* no varejo tem três aspectos diferentes:

1. *Análise da Web*. Reunir estatísticas de *website* que rastreiam o comportamento *online* do cliente:

hits, páginas, volumes de vendas e assim por diante. Isso ajuda a ajustar o *website* para atender as necessidades do cliente.

2. *Análise do cliente*. Análise do cliente acrescenta profundidade para entender as interações do cliente. Empresas reúnem dados de múltiplas fontes, incluindo interações em *websites*, dados de operação provenientes de compras *off-line*, e dados demográficos. Isso é fundamental em CRM e gerenciamento de receita, pois um melhor entendimento permite que uma empresa reúna os clientes em grupos.
3. *Otimização*. A otimização assegura grande compensação. Padrões sutis podem ser detectados e usados para otimizar as interações do cliente. Esse é a meta do CRM e do gerenciamento de receita.

Veja a J.Crew, grande varejista on-line e por catálogo de vestuário, calçados e acessórios masculinos e femininos. Ela tem tido imenso sucesso com análise de otimização. Anteriormente, a empresa usava um procedimento manual incômodo

para sugerir estilos semelhantes e complementares para compradores virtuais. No segundo semestre de 2002, a J.Crew implantou a análise de otimização. As sugestões do mecanismo analítico, que são feitas automaticamente, geram o dobro de vendas que o sistema manual antigo.

Fontes: adaptado de U. Fayyad, “Optimizing Customer Insight”, *Intelligent Enterprise*, Vol. 6 No. 8, May 1, 2003, pp. 23–26, 33; and U. Fayyad, *Optimize Customer Interactions – and Profits – with Advanced Data Mining Techniques*, watts-associates.com/docs/articles/digiMine.pdf (acessado em abril de 2006).

Aplicações de data mining

O *data mining* pode ser muito útil, como mostrado pelos seguintes exemplos representativos. O objetivo da maioria desses exemplos é identificar uma oportunidade de negócios e criar uma vantagem competitiva sustentável:

- *Marketing*. Prever quais clientes serão suscetíveis aos *banners* da Internet ou quais comprarão um produto específico e segmentar a faixa demográfica dos clientes.
- *Banco*. Prever níveis de crédito desfavoráveis e uso fraudulento de cartão de crédito, gasto no cartão de crédito de clientes novos, e quais tipos de clientes serão mais suscetíveis a uma nova oferta de crédito ou outros produtos e serviços.
- *Varejo e vendas*. Prever vendas e determinar níveis certos de estoque e distribuição de cronogramas entre as lojas.
- *Fabricação e produção*. Prever quando esperar falhas no maquinário e encontrar fatores importantes que controlam a otimização da capacidade de fabricação.
- *Corretagem e compra e venda de títulos*. Prever quando os preços dos títulos mudarão, prever a variação da flutuação de ações para assuntos específicos e para o mercado global, e determinar quando negociar as ações.
- *Seguros*. Prever quantidades de ações e custos de cobertura médica, classificar os elementos mais importantes que afetam a cobertura médica e prever quais clientes comprarão novas apólices com itens especiais.
- *Hardware e software para computador*. Prever falha na unidade de disco, prever quanto tempo levará para criar novos *chips* e prever possíveis violações de segurança.
- *Governo e defesa*. Prever o custo de mudança do equipamento militar, testar estratégias para ações militares e prever consumo de recursos.
- *Empresas aéreas*. Captura de dados não apenas de onde os clientes voam, mas também do último destino dos passageiros que mudam de empresa no meio do itinerário. Com essa informação, as empresas aéreas podem identificar locais comuns onde não prestam serviços atualmente a fim de adicionar rotas e conquistar negócios perdidos.
- *Saúde*. Correlacionar a faixa demográfica dos pacientes com doenças graves. Usando o *data mining*, os médicos podem desenvolver melhores *insights* sobre sintomas e como fornecer tratamentos adequados. Veja o Caso de aplicação 4.4 para um uso emergente das tecnologias de *data mining*.
- *Transmissão*. Prever quais programas têm maior audiência durante o horário nobre e como maximizar o lucro ao saber onde inserir propagandas.
- *Polícia*. Rastrear padrões de crime, localizações, comportamento criminoso e características para ajudar a solucionar processos penais.
- *Segurança nacional*. Amplamente usado em atividades de segurança nacional. Veja o Caso de aplicação 4.5 para um exemplo recente do uso de *data mining* para rastrear financiamento de atividades terroristas.

CASO DE APLICAÇÃO 4.4

A personalização da medicina

A pesquisa tem percorrido um longo caminho na personalização da saúde. Nos tratamentos clínicos, os médicos estão tentando encontrar a melhor maneira de personalizar o tratamento (p.ex., escolher o tratamento quimioterápico mais apropriado para um paciente com câncer com um marcador genético específico). Para uma melhor personalização do tratamento, a Mayo Clinic e a IBM estão aplicando reconhecimento de padrões e *data mining* aos registros dos pacientes.

Os dados são coletados de diferentes recursos nos hospitais da Mayo, incluindo arquivos digitais dos pacientes, resultados laboratoriais, raios-x e eletrocardiogramas. Algoritmos personalizados são aplicados nos dados para identificar padrões baseados na idade, histórico médico, genética e outros fatores. A análise relata como os pacientes responderiam a vários tratamentos e como personalizar seu cuidado. Isso também ajuda os pesquisadores no desenvolvimento de terapias. Ferramentas de reconhecimento de padrões são usadas para descobrir relações entre

proteínas, código genético e respostas a tratamentos específicos. As tecnologias de *data mining* são extremamente úteis na descoberta de correlações e revelações que anteriormente teriam sido ignoradas.

Mudar para registros clínicos eletrônicos traz consigo o problema constante de responsabilidade dos dados. Somente pessoas essenciais devem ter acesso às informações do paciente, e o *data mining* precisa ser aplicado apenas em dados agregados. Na Mayo Clinic, os pacientes podem decidir se seus dados clínicos podem ser usados para análise. A Mayo nomeou seu projeto de *data mining* de Data Trust, para lembrar aos funcionários que trabalham no projeto que o sucesso depende da privacidade, segurança e sigilo dos dados do paciente.

Fontes: M.K. McGee, "Mayo Builds Toward Customized Medicine", *InformationWeek*, August 9, 2004, p. 24; and Mayo Clinic, IBM Aim to Drive Medical Breakthroughs, mayoclinic.org/spotlight/mayoibmcollaboration.html (acessado em abril de 2006).

CASO DE APLICAÇÃO 4.5

Uma mina no financiamento de terroristas

O ataque terrorista ao World Trade Center, em 11 de setembro de 2001, realçou a importância de uma inteligência de fontes abertas. A Lei Patriota dos EUA e a criação do Departamento Norte-Americano de Segurança Nacional (DHS) anunciaram a possível aplicação de tecnologia da informação e técnicas de *data mining* para detectar lavagem de dinheiro e outras formas de financiamento terrorista. Os órgãos policiais têm se concentrado na investigação das atividades de lavagem de dinheiro através de operações normais por bancos e outras empresas de serviços financeiros. Os preços no comércio internacional vieram à baila desde que os ataques terroristas viraram o foco dos órgãos policiais.

O comércio internacional tem sido usado pelos lavadores para mover dinheiro silenciosamente para fora de um país sem atrair a atenção do governo. Supervalorizar as importações e desvalorizar as exportações são estratégias para realizar essa transferência. O modo de agir na supervalorização pode ser a seguinte: um importador nacional e um exportador estrangeiro podem formar um vínculo e supervalorizar as importações, com isso transferindo dinheiro do país de

origem, resultando em crimes relacionados a fraude aduaneira, evasão de divisas e lavagem de dinheiro. O exportador estrangeiro pode ser membro de uma organização terrorista.

As técnicas de *data mining* concentram-se na análise de dados nas operações de importação e exportação do Departamento de Comércio dos EUA e de outros organismos relacionados ao comércio. Os preços de importação que excederem os preços de importação do quartil mais alto e os preços de exportação que forem inferiores aos preços de exportação do quartil mais baixo são rastreados. O foco é nos preços irregulares de transferência entre corporações que podem resultar no deslocamento de rendas e impostos taxáveis para fora dos Estados Unidos. Uma variação de preço observada pode estar relacionado a sonegação fiscal/evasão de divisas, lavagem de dinheiro ou financiamento terrorista. A variação de preço observada também pode ser devido a um erro no banco de dados de comércio dos EUA.

O total de dinheiro estimado levado para fora dos EUA em 2001 foi de US\$156,22 bilhões. O *data mining*

resultará em avaliação eficiente de dados, que por sua vez auxiliará na cruzada contra o terrorismo. A aplicação de tecnologia da informação e técnicas de data mining às operações financeiras pode contribuir com o aumento na qualidade da informação do serviço secreto.

Fontes: J.S. Zdanowic, “Detecting Money Laundering and Terrorist Financing via Data Mining”, *Communications of the ACM*, Vol. 47, No. 5, May 2004, p. 53; and R.J. Bolton, “Statistical Fraud Detection: A Review”, *Statistical Science*, Vol. 17, No. 3, January 2002, p. 235.

Palshikar (2001) forneceu diversos exemplos de *data mining* eficaz na prática. Veja o Caso de aplicação 4.7, ainda neste capítulo, para informações sobre iniciativas de *data mining* e de análise no DHS. Dados do censo podem ser combinados com outros dados de mercado na segmentação de clientes (veja Grimes, 2001).

Uma aplicação menos característica de *data mining* foi usada para melhorar o desempenho dos times da NBA (National Basketball Association), nos Estados Unidos. A NBA desenvolveu o Advanced Scout, uma aplicação de *data mining* baseada em PC, que a equipe de treinamento usa para descobrir padrões interessantes nos dados dos jogos de basquete. O processo de interpretação de padrões é facilitado ao permitir que o usuário relacione padrões ao videoteipe. Para obter mais detalhes, veja Bhandari et al. (1997).

Quando as empresas são assoladas por fraudes, especialmente em operações financeiras, como no comércio eletrônico, elas aplicam ferramentas de *data mining* especializadas para detectar padrões nos dados. Em geral, esses métodos usam redes neurais em acréscimo aos métodos estatísticos e de agrupamento. O SAS Anti-Money Laundering é um exemplo de *software* que pode fazer isso. Veja o caso de sucesso do Unity Trust Bank em sas.com/news/sascom/2003q3/feature_launderer.html.

Uma equipe de biólogos noruegueses desenvolveu métodos inteligentes para procurar na e explorar a Web em busca de estudos genéticos que contenham informações relevantes para seus propósitos. Métodos como esses tornam-se cada vez mais importantes para pesquisadores científicos e também para os executivos, pois a quantidade de informação gerada e armazenada dobra a cada três anos (veja Pallatto, 2002). Veja Copeland (2001) para saber como o *Web mining* é executado.

Questões de revisão da Seção 4.2

1. Defina *data mining*.
2. Quais fatores aumentaram a popularidade do *data mining* recentemente?
3. Identifique pelo menos cinco aplicações de *data mining*.
4. Quais são as principais características do *data mining*?
5. Quais são algumas das principais categorias das tecnologias de *data mining*? Defina pelo menos três delas.

4.3

TÉCNICAS E FERRAMENTAS DE DATA MINING

Existem muitos métodos para realização de *data mining*. Um *software* de *data mining* pode usar uma ou mais dessas técnicas. Ferramentas e técnicas de *data mining* podem ser classificadas com base na estrutura dos dados e algoritmos usados. Essas são as principais:

- *Métodos estatísticos*. Eles incluem regressão linear e não-linear, estimativa de ponto, teorema de Bayes (ou seja, distribuição de probabilidade), correlações e análise de *cluster*.
- *Árvores de decisão*. As **árvores de decisão** são usadas em métodos de classificação e agrupamento. Elas decompõem os problemas em subconjuntos progressivamente discretos ao trabalhar de generalizações a informações cada vez mais específicas. Uma árvore de decisão pode ser definida como uma raiz seguida de nós internos. Cada nó (incluindo a raiz) é rotulado

com uma questão. Os arcos associados a cada nó abrangem todas as respostas possíveis. Cada resposta representa um resultado provável (veja Dunham, 2003).

- *Raciocínio baseado em casos.* Essa abordagem usa casos históricos para reconhecer padrões. Por exemplo, os clientes da Cognitive Systems Inc. usam essa abordagem para auxiliar nas aplicações de *helpdesk*. Um cliente tem uma biblioteca com 50.000 casos de consulta. Casos novos podem ser rapidamente combinados com as 50.000 amostras na biblioteca, e as respostas são fornecidas automaticamente às consultas, com mais de 90% de precisão.
- *Computação neural.* Redes neurais usam muitos nós relacionados (que funcionam de maneira semelhante aos neurônios no cérebro humano). Essa abordagem examina uma quantidade volumosa de dados históricos em busca de padrões. Assim, podemos percorrer grandes bancos de dados e, por exemplo, identificar possíveis clientes para um produto novo ou empresas cujos perfis sugerem que estão indo à falência. Aplicações de computação neural são comuns em serviços financeiros (veja Fadlalla e Lin, 2001) e fabricação. Uma descrição ampla das redes neurais é fornecida no Capítulo online 6.
- *Agentes inteligentes.* Uma das abordagens mais promissoras para recuperar informações de bancos de dados, especialmente os externos, é o uso de agentes inteligentes. Com a disponibilidade de uma vasta e crescente quantidade de informações pela Internet, encontrar a informação certa está se tornando cada vez mais difícil. Aplicações de *data mining* baseadas na Web normalmente são ativadas por agentes de *software* inteligentes.
- *Algoritmos genéticos.* Algoritmos genéticos baseiam-se no princípio de expansão de resultados possíveis. Dado um número fixo de resultados possíveis, os algoritmos genéticos buscam definir soluções novas e melhores. Eles são usados para regras de agrupamento e associação.
- *Outras ferramentas.* Muitas outras ferramentas podem ser usadas para *data mining*, incluindo indução de regras e visualização de dados. A melhor fonte de informações sobre o desenvolvimento de novas ferramentas é o *site* do fornecedor.

Ao lidar com dados comportamentais do cliente, que podem conter 100 ou mais dimensões, os algoritmos são capazes de lidar com dados de alta dimensão de maneira efetiva. Esses algoritmos também devem trabalhar com regras e restrições comerciais. A estatística simples não trabalha. Conhecimento das restrições do negócio, das relações entre produtos e dos vários segmentos comportamentais dos clientes é uma necessidade. As próximas seções apresentam algoritmos para dois problemas comuns abordados pelo *data mining*: classificação e agrupamento.

Classificação

Classificação, como o nome indica, envolve a identificação de padrões de dados para colocá-los em determinadas categorias. Essas tarefas podem incluir aprovação de crédito (ou seja, risco de crédito bom ou ruim), localização de loja (p.ex., boa, moderada, ruim), *marketing* direcionado (p.ex., cliente propenso, sem chance), detecção de fraude (ou seja, sim ou não) e telecomunicação (p.ex., propenso a mudar para outra empresa telefônica, sim/não). Esses métodos envolvem semear um conjunto de dados com um conjunto de classes conhecido e mapear todos os outros itens (p.ex., clientes) naqueles conjuntos. A classificação pode ser baseada em classes previamente conhecidas ou totalmente desconhecidas. Quando as classes não são conhecidas para nenhum dado, a *análise de cluster* é usada.

Uma aplicação mais comum da classificação é desenvolver um modelo que aprenda a partir de classes de padrões de dados previamente conhecidas e, em seguida, seja capaz de classificar novos dados dentro das classes. Então, as entradas no modelo são um conjunto de padrões de treinamento, cada um com um rótulo de classe. O modelo atribui um rótulo de classe para cada padrão, com base na combinação de características. Após um modelo ser treinado, pode ser usado para prever classes de novos padrões nos quais o rótulo seja desconhecido. O processo básico de criação desses modelos é treinar um algoritmo usando dados amostrais cujas classes sejam conhecidas. O desempenho do algoritmo é então testado ao receber um conjunto de dados no

qual a classe atual é conhecida, mas é pedido ao modelo a sua previsão. Os algoritmos podem ser comparados com base em uma série de fatores:

- *Acurácia preditiva*, que se refere à capacidade de o modelo prever corretamente o rótulo de classe de um dado novo ou previamente despercebido. Para calcular essa precisão, uma classe de amostras de teste conhecida é combinada com a classe predita pelo modelo. Então, a precisão pode ser calculada como Taxa de precisão = Porcentagem de amostras de teste classificadas corretamente pelo modelo. Naturalmente, dependendo do problema, talvez também estejamos interessados em estimar os erros do modelo em relação à classificação inadequada específica.
- *Velocidade*, que se refere aos custos computacionais envolvidos na geração e uso do modelo.
- *Robustez*, que é a capacidade do modelo fazer previsões certas, dado ruídos ou dados sem valores.
- *Escalabilidade*, que se refere à habilidade de criar modelos de modo eficaz, dado uma grande quantidade de dados.
- *Interpretabilidade*, que se refere ao nível de entendimento e *insight* fornecido pelo modelo.

Existem muitas técnicas de classificação em uso atualmente, incluindo:

- Classificação por árvore de decisão
- Análise estatística
- Redes neurais
- Classificador Bayesiano
- Algoritmos genéticos
- Teoria dos conjuntos aproximados

Uma descrição completa de todas essas técnicas de classificação está fora do escopo deste livro. Redes neurais são abordadas no Capítulo 6. A seguir, nos concentraremos na classe de algoritmos chamada árvore de decisão.

Árvores de decisão

Para descrever uma árvore de decisão, precisamos introduzir algumas terminologias simples. Primeiro, os padrões de dados incluem muitas variáveis de entrada que podem ter um impacto na classificação desses padrões. Essas variáveis de entrada normalmente são chamadas de *atributos*. Por exemplo, se criarmos um modelo para classificar riscos de crédito com base em apenas duas características – renda e índice de crédito –, elas seriam os atributos. A informação resultante seria o rótulo de classe. Segundo, uma árvore consiste de ramificações e nós. Uma ramificação representa o resultado de um teste em classificar um padrão com base em um teste, usando um atributo. Um nó folha no fim representa a opção de classe final para um padrão. Um nó intermediário representa um teste em um atributo.

A ideia básica de uma árvore de decisão é que ela divide, de maneira recorrente, um conjunto de treinamento até que cada divisão consista inteira ou primariamente de exemplos provenientes de uma única classe. Cada nó não-folha da árvore contém um *ponto de divisão*, que é um teste em um ou mais atributos e determina como os dados serão divididos posteriormente. Em geral, algoritmos de árvore de decisão criam uma árvore inicial, a partir dos dados de treinamento, para que cada nó folha seja puro e então eles podam a árvore para aumentar sua precisão nos dados de teste.

Na fase de crescimento, a árvore é criada pela divisão repetitiva dos dados até que cada divisão também seja pura (ou seja, contenha membros da mesma classe) ou relativamente pequena. A ideia básica é fazer perguntas cujas respostas forneçam a maioria da informação, semelhante ao que podemos fazer quando jogamos o 20Q.

A divisão usada para separar os dados depende do tipo de atributo usado na divisão. Para um atributo contínuo A , as divisões são do valor da forma $(A) < x$, onde x é um valor de A . Por exemplo, a divisão baseada na renda poderia ser “Renda < 50.000”. Para um atributo categórico A , as divisões são do valor da forma (A) que pertence a X , onde X é um subconjunto de A . Como exemplo, a divisão poderia ser baseada no gênero: “O gênero é masculino”.

O algoritmo mais usual para criação de uma árvore de decisão é:

1. Criar um nó raiz e selecionar um atributo de divisão.
2. Adicionar uma ramificação ao nó raiz para cada valor e rótulo candidato a divisão.
3. Seguir as seguintes etapas repetidas:
 - a. Classificar os dados através da aplicação do valor de divisão.
 - b. Se um ponto de parada for atingido, criar um nó folha e rotulá-lo. Caso contrário, criar outra subárvore.

Muitos algoritmos diferentes foram propostos para criação de árvores de decisão. Os algoritmos diferem principalmente em relação a escolha de atributos de divisão, ordem dos atributos de divisão, número de divisões, estrutura da árvore, critério de parada e poda da árvore. Alguns dos algoritmos mais famosos são ID3, C4.5, C5 do aprendizado de máquina, CART (Classification and Regression Trees) da estatística e CHAID (Chi-squared Automatic Interaction Detector) do reconhecimento de padrão.

Durante a criação de uma árvore de decisão, o objetivo em cada nó é determinar o ponto de divisão que melhor separa os registros de treinamento pertencentes àquela folha. Para avaliar a qualidade da divisão, alguns índices de divisão foram propostos. Dois dos mais comuns são o índice Gini e o ganho de informação. O índice Gini é usado em algoritmos CART e SPRINT (Scalable Parallelizable Induction of Decision Trees). Versões do ganho de informação são usados no ID3, C4.5 e suas versões novas. Descrevemos o índice Gini a seguir para ilustrar o que acontece na criação de uma árvore de decisão.

O índice Gini O índice Gini tem sido usado na economia para medir a diversidade da população. O mesmo conceito pode ser usado para determinar a pureza de uma classe específica como resultado de uma decisão de ramificar mais um atributo ou variável específica. A melhor divisão é aquela que aumenta a pureza dos conjuntos resultantes de uma divisão proposta. Considere, por exemplo, um problema simples na criação de uma árvore de decisão para os dados na Tabela 4.2. Essa tabela possui três colunas para relacionar renda e índice de crédito com uma avaliação do risco de crédito. Se pudéssemos criar um classificador baseado nos dados históricos fornecidos pela Tabela 4.2, poderíamos usá-lo para avaliar os riscos de crédito para outras aplicações futuras. (Obviamente, sabemos que a aplicação na vida real provavelmente envolveria mais do que esses dois fatores, mas queremos mantê-la simples para podermos aprender os detalhes do algoritmo.) A segunda coluna da Tabela 4.2 fornece a renda, a terceira fornece o índice de crédito e a quarta lista o risco de crédito correspondente.

Primeiro, vamos definir o índice Gini. Ele é calculado como:

$$\text{Gini}(S) = 1 - \sum p_j^2$$

onde:

S é um conjunto de dados que contém exemplos de n classes

p_j é uma frequência relativa de classe j em S

Se, por exemplo, existem duas classes, Alto e Baixo, e o conjunto de dados S com elementos p Alto e elementos n Baixo, usamos:

$$p_{\text{alto}} = p/(p + n)$$

$$p_{\text{baixo}} = n/(n + p)$$

$$\text{Gini}(S) = 1 - p_{\text{alto}}^2 - p_{\text{baixo}}^2$$

Tabela 4.2 Padrões de amostra para treinamento de uma árvore de decisão a fim de prever risco de crédito

<i>Padrão n°</i>	<i>Renda</i>	<i>Índice de crédito</i>	<i>Risco de crédito</i>
0	23	Alto	Alto
1	17	Baixo	Alto
2	43	Baixo	Alto
3	68	Alto	Baixo
4	32	Moderado	Baixo
5	20	Alto	Alto

Se o conjunto de dados S é dividido em S_1 e S_2 , o índice de divisão é definido como:

$$\text{Gini}_{\text{DIVISÃO}}(S) = (p_1 + n_1)/(p + n) \times \text{Gini}(S_1) + (p_2 + n_2)/(p + n) \times \text{Gini}(S_2)$$

onde p_1, n_1 (p_2, n_2) indicam elementos p_1 Alto e elementos n_1 Baixo no conjunto de dados S_1 (S_2). Nessa definição, o melhor ponto de divisão é aquele com o menor valor do índice $\text{Gini}_{\text{DIVISÃO}}$.

Para ilustrar o uso do índice Gini no desenvolvimento de uma árvore de decisão, considere os dados da Tabela 4.2. Precisamos decidir qual atributo levar em consideração primeiro para ramificação. Existem muitas implementações especiais para ajudar a tomar essa decisão, mas podemos simplesmente começar com o único atributo numérico presente nos dados: renda. Isto é, podemos tentar segmentar os dados de modo que a escolha da renda possa nos ajudar a determinar se um risco de crédito é alto ou baixo. Podemos reordenar os dados da Tabela 4.2 para ver o seguinte resumo em relação ao atributo Renda:

<i>Renda</i>	<i>Padrão n°</i>	<i>Risco de crédito</i>
17	1	Alto
20	5	Alto
23	0	Alto
32	4	Baixo
43	2	Alto
68	3	Baixo

Dessa maneira, os valores possíveis de um ponto de divisão para o atributo Renda são $Renda \leq 17$, $Renda \leq 20$, $Renda \leq 23$, $Renda \leq 32$, $Renda \leq 43$ e $Renda \leq 68$. Agora, podemos calcular o índice Gini para cada um desses níveis de divisão.

Considere a opção de divisão de dados em $Renda \leq 17$. Temos as seguintes opções de classificação:

<i>Contagem padrão</i>	<i>Alto</i>	<i>Baixo</i>
$Renda \leq 17$	1	0
$Renda > 17$	3	2

Vamos calcular o índice Gini para $Renda \leq 17$ e $Renda > 17$. Primeiro,

$$G(Renda \leq 17) = 1 - (\text{Proporção de registros com risco Alto})^2 - (\text{Proporção de registros com risco Baixo})^2$$

Isto é,

$$G(Renda \leq 17) = 1 - (1^2 + 0^2) = 0$$

De forma semelhante,

$$G(\text{Renda} > 17) = 1 - ((3/5)^2 + (2/5)^2) = 1 - (13/25) = 12/25$$

O índice Gini para opção de divisão é calculado da seguinte maneira:

$$G_{\text{DIVISÃO}} = (\text{Proporção de registros com Renda} \leq 17) \times G(\text{Renda} \leq 17) + (\text{Proporção de registros com Renda} > 17) \times G(\text{Renda} > 17)$$

$$G_{\text{DIVISÃO}} = (1/6) \times 0 + (5/6) \times (12/25) = 2/5$$

Agora, considere a opção Renda ≤ 20 :

<i>Contagem padrão</i>	<i>Alto</i>	<i>Baixo</i>
Renda ≤ 20	2	0
Renda > 20	2	2

O índice Gini para essa divisão é calculado da seguinte maneira:

$$G(\text{Renda} \leq 20) = 1 - (1^2 + 0^2) = 0$$

$$G(\text{Renda} > 20) = 1 - ((1/2)^2 + (1/2)^2) = 1/2$$

$$G_{\text{DIVISÃO}} = (2/6) \times 0 + (4/6) \times (1/2) = 1/3$$

De forma semelhante, considere divisão Renda em 23:

<i>Contagem padrão</i>	<i>Alto</i>	<i>Baixo</i>
Renda ≤ 23	3	0
Renda > 23	1	2

O índice Gini para essa divisão é calculado da seguinte maneira:

$$G(\text{Renda} \leq 23) = 1 - (1^2 + 0^2) = 0$$

$$G(\text{Renda} > 23) = 1 - ((1/3)^2 + (2/3)^2) = 1 - (1/9) - (4/9) = 4/9$$

$$G_{\text{DIVISÃO}} = (3/6) \times 0 + (3/6) \times (4/9) = 2/9$$

Por fim, considere a divisão Renda em 32:

<i>Contagem padrão</i>	<i>Alto</i>	<i>Baixo</i>
Renda ≤ 32	3	1
Renda > 32	1	1

O índice Gini para essa divisão é calculado da seguinte maneira:

$$G(\text{Renda} \leq 32) = 1 - ((3/4)^2 + (1/4)^2) = 1 - (10/16) = 6/16 = 3/8$$

$$G(\text{Renda} > 32) = 1 - ((1/2)^2 + (1/2)^2) = 1/2$$

$$G_{\text{DIVISÃO}} = (4/6) \times (3/8) + (2/6) \times (1/2) = (1/8) + (1/6) = 14/48 = 7/24$$

O menor valor de $G_{\text{DIVISÃO}}$ é para Renda ≤ 23 . Então, pegamos os dois valores mais próximos e calculamos a média. Assim, temos o ponto de divisão em Renda = $(23 + 32)/2 = 27,5$.

As listas de atributos são separadas no ponto de divisão. Isto é, esperamos ter uma regra que diga:

Se Renda $\leq 27,5$

Então...

Senão Se Renda > 27,5
Então...

Agora, examinamos os padrões para preencher a parte Então da regra. Vamos ver se podemos abranger todos os registros de uma categoria específica ao mesmo tempo.

A seguir está a lista de atributos para Renda ≤ 27,5:

Renda	Padrão n°	Risco de crédito
17	1	Alto
20	5	Alto
23	0	Alto

Ao examinar esses mesmos registros para Índice de crédito, revelamos também que:

Índice de crédito	Padrão n°	Risco de crédito
Moderado	0	Alto
Baixo	1	Alto
Moderado	5	Alto

Portanto, parece que uma regra pode ser formulada: Se Renda ≤ 27,5, então Risco de crédito = Alto, independentemente do nível de Índice de crédito.

Agora, vamos examinar a outra ramificação dessa árvore de decisão através da verificação das listas de atributos para Renda > 27,5. A tabela a seguir sugere que Renda > 27,5 não é um indicador decisivo de Risco de crédito:

Renda	Padrão n°	Risco de crédito
32	4	Baixo
43	2	Alto
68	3	Baixo

Índice de crédito	Padrão n°	Risco de crédito
Baixo	2	Alto
Moderado	3	Baixo
Alto	4	Baixo

Agora, desenvolvemos uma subárvore através da verificação do atributo Índice de crédito.

Índice de crédito é uma variável categórica. As regras para o desenvolvimento de uma ramificação de decisão para uma variável categórica, como gênero, são levemente diferentes daquelas para um atributo contínuo com valor real, como renda. Lembre-se que o índice Gini está definido como:

$$\text{Gini}(\text{duas proporções}) = 1 - p_{\text{uma_proporção}}^2 - p_{\text{a_outra_proporção}}^2$$

No caso de uma variável categórica, isso significa que, por exemplo, uma proporção é o conjunto de registros de Índice de crédito = {Baixo}. A outra proporção significa, então, Índice de crédito = não {Baixo}. No nosso caso, isso também significa: Índice de crédito ∈ {Moderado, Alto}. Portanto, temos de calcular as proporções de cada categoria e seu complemento. Em geral, precisamos avaliar o índice de divisão para cada uma das combinações 2^N , onde N é a cardinalidade do atributo categórico:

<i>Contagem padrão</i>	<i>Alto</i>	<i>Baixo</i>
Índice de crédito = {Baixo}	1	0
Índice de crédito = {Moderado}	0	1
Índice de crédito = {Alto}	0	1

Primeiro, calculamos o índice Gini para cada categoria:

$$\begin{aligned} G(\text{Índice de crédito} = \{\text{Baixo}\}) &= 1 - 1^2 - 0^2 = 0 \\ G(\text{Índice de crédito} = \{\text{Moderado}\}) &= 1 - 0^2 - 1^2 = 0 \\ G(\text{Índice de crédito} = \{\text{Alto}\}) &= 1 - 0^2 - 1^2 = 0 \end{aligned}$$

Isso não é uma surpresa, pois existe somente um registro de cada tipo. A seguir, calculamos o índice Gini para as categorias complementares:

$$\begin{aligned} G(\text{Índice de crédito} \in \{\text{Baixo, Moderado}\}) &= 1 - (1/2)^2 - (1/2)^2 = 1/2 \\ G(\text{Índice de crédito} \in \{\text{Baixo, Alto}\}) &= 1/2 \\ G(\text{Índice de crédito} \in \{\text{Moderado, Alto}\}) &= 1 - 0^2 - 1^2 = 0 \end{aligned}$$

Por fim, podemos calcular os valores do índice Gini para as ramificações possíveis. Para uma opção de ramificação de Índice de crédito = Baixo e Índice de crédito = {Moderado, Alto}, teríamos:

$$\begin{aligned} G_{\text{DIVISÃO}} &= (\text{Proporção de registros com Índice de crédito} = \text{Baixo}) \\ &\times G(\text{Índice de crédito} \in \{\text{Baixo}\}) + (\text{Proporção de registros com} \\ &\text{Índice de crédito} = \text{não Baixo}) \times G(\text{Índice de crédito} = \text{não } \{\text{Baixo}\}) \end{aligned}$$

Ou, poderíamos ter:

$$\begin{aligned} G_{\text{DIVISÃO}} &= (\text{Proporção de registros com Índice de crédito} = \text{Baixo}) \\ &\times G(\text{Índice de crédito} \in \{\text{Baixo}\}) + \\ &(\text{Proporção de registros com Índice de crédito} = \text{Moderado, Alto}) \\ &\times G(\text{Índice de crédito} = \text{Moderado, Alto}) \end{aligned}$$

Isto é:

$$G_{\text{DIVISÃO}}(\text{Índice de crédito} = \{\text{Baixo}\}) = (1/3) \times 0 + (2/3) \times 0 = 0$$

O restante dos índices Gini para outras opções de divisão podem ser calculadas da seguinte maneira:

$$\begin{aligned} G_{\text{DIVISÃO}}(\text{Índice de crédito} = \{\text{Moderado}\}) &= (1/3) \times 0 + (2/3) \times (1/2) = 1/3 \\ G_{\text{DIVISÃO}}(\text{Índice de crédito} = \{\text{Alto}\}) &= (1/3) \times 0 + (2/3) \times (1/2) = 1/3 \\ G_{\text{DIVISÃO}}(\text{Índice de crédito} \in \{\text{Baixo, Moderado}\}) &= (2/3) \times (1/2) + (1/3) \times 0 = 1/3 \\ G_{\text{DIVISÃO}}(\text{Índice de crédito} \in \{\text{Baixo, Alto}\}) &= (2/3) \times (1/2) + (1/3) \times 0 = 1/3 \\ G_{\text{DIVISÃO}}(\text{Índice de crédito} \in \{\text{Moderado, Alto}\}) &= (2/3) \times 0 + (1/3) \times 0 = 0 \end{aligned}$$

O menor valor do índice Gini para a divisão é zero em Índice de crédito = Baixo e Índice de crédito ∈ {Moderado, Alto}, então esse é nosso ponto de divisão, e essas são as ramificações do próximo nível. A árvore de decisão que nós criamos é mostrada na Figura 4.1.

Ganho de informação O índice Gini é apenas um dos muitos métodos para criar uma árvore de decisão. Talvez a abordagem de árvore de decisão mais amplamente conhecida envolva o algoritmo ID3 (veja Quinlan, 1986) e suas variantes. A idéia básica do algoritmo ID3 é usar um conceito chamado *entropia* em vez do índice Gini. A **entropia** mede o grau de incerteza ou aleatoriedade em um conjunto de dados. Se todos os dados em um subconjunto pertencerem a apenas uma classe, não há incerteza ou aleatoriedade naquele conjunto de dados; a entropia é zero. O objetivo dessa abordagem é criar subárvores de modo que a entropia de cada subconjunto final seja zero.

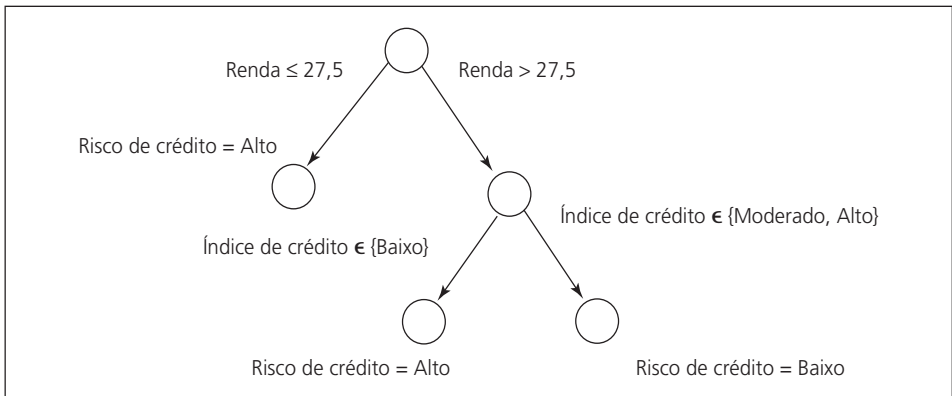


Figura 4.1 Árvore de decisão usando o índice Gini para o critério de divisão.

Como é o caso dos algoritmos CART/SPRINT que usam o índice Gini, a entropia de cada subconjunto candidato é calculada, e então uma medida de ganho de informação é calculada para determinar qual será o modo de divisão. As bases são as mesmas; os detalhes específicos dos algoritmos, naturalmente, variam. Quinlan (1986) descreveu o algoritmo ID3 básico. A empresa de Quinlan oferece versões mais recentes do ID3, chamadas C4.5 e See5. A seção “*Software de data mining*”, ainda neste capítulo, descreve muitas ofertas relacionadas.

Análise de *cluster* para data mining

*Análise de cluster** é um conjunto de métodos muito importante para classificação de itens em grupos comuns chamados *clusters*. Estes métodos são populares em biologia, medicina, genética, ciência social, antropologia, arqueologia, astronomia, reconhecimento de caracteres e até mesmo no desenvolvimento de sistemas de informações gerenciais (SIG). Com o aumento da popularidade do *data mining*, os métodos têm sido aplicados aos negócios, em especial ao *marketing*. A análise de *cluster* têm sido muito usada para detecção de fraude, tanto de cartão de crédito quanto de comércio eletrônico, e segmentação de mercado de clientes em sistemas de CRM. Mais aplicações em negócios continuam a ser desenvolvidas à medida que a força da análise de *cluster* é compreendida e usada.

Análise de *cluster* é uma ferramenta de análise exploratória de dados para a solução de problemas de classificação. O objetivo é ordenar casos (p.ex., pessoas, coisas, eventos) em grupos, ou *clusters*, de modo que o grau de associação seja forte entre os membros do mesmo *cluster* e fraco entre membros de *clusters* diferentes. Cada *cluster* descreve a classe que seus membros pertencem. Um exemplo unidimensional evidente dessa análise é estabelecer faixas de pontuação para assinalar notas de aula para uma turma de faculdade. Esse problema é semelhante ao de análise de *cluster* enfrentado pelo Departamento do Tesouro dos EUA ao estabelecer novas faixas de impostos nos anos 80. Um exemplo fictício de agrupamento acontece no livro Harry Potter, de J.K. Rowling. O Chapéu Seletor determina para qual Casa (p.ex., dormitório) designar os alunos do primeiro ano na Escola de Hogwarts. Um outro exemplo envolve determinar como acomodar os convidados em um casamento. Na medida em que o *data mining* avança, a importância da análise de *cluster* é revelar associações e estruturas em dados que não estavam aparentes anteriormente, mas que são sensíveis e úteis uma vez achados.

Os resultados da análise de *cluster* podem ser usados para:

- Ajudar a identificar um plano de classificação (p.ex., tipos de clientes)
- Sugerir modelos estatísticos para descrever populações

* N. de R.: Alguns autores nacionais traduzem a palavra *cluster* por conglomerados.

- Indicar regras para atribuir novos casos às classes para fins de identificação, direcionamento e diagnóstico
- Fornecer medidas de definição, tamanho, e mudar o que antes eram conceitos amplos
- Encontrar casos típicos para representar classes

Métodos de análise de *cluster*

A análise de *cluster* pode ser baseada em um ou mais dos seguintes métodos gerais:

- Métodos estatísticos (incluindo tanto hierárquico como não-hierárquico)
- Métodos ideais
- Redes neurais
- Lógica *fuzzy*
- Algoritmos genéticos

Geralmente, cada um desses métodos funciona com uma das classes de métodos gerais:

- *Divisivo*. Com classes divisoras, todos os itens começam em um *cluster* e são decompostas.
- *Aglomerativo*. Com classes aglutinadoras, todos os itens começam em *clusters* distintos, que depois são unidos.

A maioria dos métodos de análise de *cluster* envolve o uso de uma distância entre os pares de itens. Isto é, há uma medida de semelhança entre cada par de itens a ser agrupado. Muitas vezes, eles estão baseados em distâncias exatas que são medidas, mas essa necessidade não é verdadeira, como é tipicamente o caso no desenvolvimento de sistemas de informação (SI). Por exemplo, em um projeto de desenvolvimento de SI, módulos distintos do sistema podem ser relacionados pela semelhança entre suas entradas, saídas, processos e os dados específicos usados. Então, esses fatores são reunidos, em pares por item, em uma única medida de distância.

Exemplo de agrupamento

Considere a matriz de semelhança (distância) que representa as similaridades entre oito itens mostrados na Tabela 4.3. Os itens 4 e 5 têm muito em comum, assim como os itens 1 e 3 e 3 e 10; 1 e 10 estão ligeiramente relacionados, e 1 e 5 têm pouco em comum. Para avaliar uma solução, adicionamos os valores em pares de todos os itens em cada *cluster*. Se nós queremos três *clusters* equilibrados (ou seja, 2 ou 3 itens por *cluster*), a solução *clusters* {1, 3, 6}, {2, 8} e {4, 5, 7} tem um valor de $(9 + 6 + 10) + 8 + (10 + 8 + 9) = 60$. Podemos fazer melhor? Tente e verifique!

Tabela 4.3 Matriz de semelhança (distância)

Os valores abaixo da diagonal são iguais aos valores acima; isto é, a distância entre 1 e 2 é a mesma entre 2 e 1. Não existem valores diagonais.

Item	1	2	3	4	5	6	7	8
1	–	3	9	2	1	6	4	5
2		–	4	5	6	2	3	8
3			–	5	7	10	4	2
4				–	10	2	8	1
5					–	4	9	3
6						–	3	3
7							–	5
8								–

Agora que temos um conjunto de dados, precisamos abordar algumas questões críticas:

- Quantos *clusters* podem ser encontrados (ou seja, quando paramos)?
- Todos os *clusters* devem ter aproximadamente o mesmo número de itens?
- Como lidamos com escala dimensional quando medidas diferentes são usadas no estabelecimento do valor de distância?
- A distância realmente pode ser medida?

Método e exemplo de agrupamento hierárquico

Como um exemplo de agrupamento hierárquico, começamos com um conjunto de itens, cada um dentro do seu próprio *cluster*. Determinamos o número máximo de *clusters* que queremos ter. O método básico é o seguinte:

1. Decidir quais dados provenientes dos itens registrar (ou seja, medidas de semelhança)
2. Calcular as distâncias entre todos os *clusters* iniciais. Armazenar os resultados em uma matriz de distância.
3. Pesquisar na matriz de distância e encontrar os dois *clusters* mais semelhantes.
4. Unificar esses dois *clusters* para produzir um outro que tenha pelo menos dois itens.
5. Calcular as distâncias entre esse novo cluster e todos os outros (alguns podem conter um único item).
6. Repetir os passos 3 ao 5 até que se tenha atingido o número máximo de *clusters* predefinido.

Observe que alguns métodos vão até o fim com um único *cluster* de todos os itens. Para identificar a solução desejada, é preciso identificar quando o número desejado foi obtido e então parar.

Por exemplo, podemos aplicar o método hierárquico à matriz da seção anterior, com um objetivo de três *clusters* equilibrados. A solução inicial é:

{1, 2, 3, 4, 5, 6, 7, 8} com um valor de 134.

Embora esse seja um valor excelente, queremos três *clusters*. Por isso, primeiro combinamos os itens 4 e 5 (valor = 10) para obter:

{4, 5} {1, 2, 3, 6, 7, 8} com $10 + 71 = 81$.

Isso é muito bom. Em seguida, combinamos os itens 3 e 6 (valor = 10) para obter:

{3, 6} {4, 5} {1, 2, 7, 8} com $10 + 10 + 28 = 48$.

Então, unimos o item 7 ao cluster com os itens 4 e 5 para obter:

{3, 6} {4, 5, 7} {1, 2, 8} em = 53.

Paramos porque temos três *clusters* tão equilibrados quanto foi possível com esses dados (ou seja, dois *clusters* de três itens, um *cluster* de dois itens).

Embora não esteja equilibrado, a solução de *cluster* {1, 3, 4}, {3, 5, 6} e {5} vale 72, que é melhor do que as soluções descritas anteriormente.

Software de data mining

Os *software* de *data mining* possuem algoritmos mais complicados para redes neurais, agrupamento, segmentação, e classificações que geralmente são mais sofisticadas do que os métodos de OLAP (veja Finlay, 2001). Muitos fornecedores de *software* oferecem ferramentas de *data mining* poderosas. Exemplos desses fornecedores são Angoss Knowledge Engineering (KnowledgeSTUDIO e KnowledgeSeeker), Cognos (uma variedade de ferramentas), Teradata, SPSS (Clementine

e outros), IBM (DB2 Intelligent Miner Scoring [IMS]), Megaputer Intelligence Inc. (PolyAnalyst) e SAS (uma variedade de ferramentas). Uma outra ferramenta poderosa é o STATISTICA Data Miner da StatSoft, que está disponível para uso com tempo limitado pelos alunos e instrutores que usam este livro. O KnowledgeSeeker da Angoss produz regras a partir de dados. Essas regras podem ser usadas em sistemas especialistas. O DB2 IMS da IBM fornece análises de *data mining* relacionais em tempo real e pontuação. Ele usa a PMML (Predictive Model Markup Language) do grupo de *data mining*, que traz o processo um passo mais próximo da automação. O PolyAnalyst inclui métodos inteligentes tanto de *data mining* quanto de *text mining*. Muitas empresas (a maioria recém-citadas) estão mudando para oferecer serviços de consultoria usando os próprios produtos em vez de apenas vender o *software*.

Buck (2000) organizou as classes das ferramentas e técnicas de *data mining* conforme elas se relacionam com tecnologias da informação e do *business intelligence* (BI). Sua taxonomia é:

- Pacotes de análises matemática e estatística
- Ferramentas de personalização para *marketing* baseado na Web
- Análise desenvolvida em plataformas de *marketing*
- Ferramentas de CRM avançadas
- Análise adicionada a outras plataformas verticais específicas do setor
- Análise adicionada às ferramentas de banco de dados (p.ex., OLAP)
- Ferramentas de *data mining* independentes

WEKA é uma coleção de algoritmos de aprendizado de máquina com código aberto para tarefas de *data mining* e inclui recursos de rede neural. O *download* do WEKA pode ser feito em **cs.waikato.ac.nz/~ml/weka**. Algumas empresas de *software* disponibilizam suas ferramentas de *data mining* e de OLAP para estudantes universitários gratuitamente ou com ótimos descontos. Verifique o *site* de cada fornecedor para obter informações sobre o assunto. A Tabela 4.4 lista alguns dos principais produtos e seus respectivos *sites*.

Software de análise de cluster

À parte dos métodos de *data mining* nos quais os métodos de análise de *cluster* estão integrados, existem vários pacotes especializados para essa análise. Eles incluem ClustanGraphics (Clustan),

Tabela 4.4 Softwares de *data mining* selecionados

Nome do produto	URL
Clementine	spss.com/clementine/
Delta Master	bissantz.de
iData Analyzer	infoacumen.com
Fair Isaac Business Science	fairisaac.com/edm
GhostMiner	fq.pl/ghostminer
IBM DB2 Data Warehouse Edition	ibm.com/software/data/db2/dwe
Knowledge Miner	knowledgeminer.net
Oracle Data Mining (ODM)	otn.oracle.com/products/bi/9idmining.html
PolyAnalyst	megaputer.com/products/pa/index.php3
Salford Systems Data Mining Suite	salfordsystems.com
SAS Enterprise Miner	sas.com/technologies/bi/analytics/index.html
STATISTICA	statsoft.com/products/dataminer.htm
Teradata Warehouse Miner	ncr.com/products/software/teradata_mining.htm
XLMiner	xlminer.net

DecisionWORKS Suite (Advanced Software Applications), SPSS (SPSS) e PolyAnalyst Cluster Engine (Megaputer). Além disso, o código de análise de *cluster* grátis está disponível em muitos *sites* universitários. Faça uma pesquisa na Web para encontrá-los.

Questões de revisão da Seção 4.3

1. Identifique pelo menos três dos principais métodos de *data mining*.
2. Dê exemplos de situações nas quais a classificação seria uma técnica de *data mining* adequada.
3. Liste pelo menos duas técnicas de classificação.
4. Quais são alguns dos critérios para comparação e seleção da melhor técnica de classificação?
5. Faça uma breve descrição do algoritmo geral usado nas árvores de decisão.
6. Defina *índice Gini*. O que ele avalia?
7. Dê exemplos de situações nas quais a análise de *cluster* seria uma técnica de *data mining* adequada.
8. Qual é a principal diferença entre análise de *cluster* e classificação?
9. Quais são alguns dos métodos para análise de *cluster*?
10. Qual é a medida de semelhança em um ambiente de análise de *cluster*?

4.4

PROCESSOS DO PROJETO DE DATA MINING

De forma semelhante às outras iniciativas importantes em uma empresa, os projetos de *data mining* precisam seguir um processo de gerenciamento de projeto. Adeptos e pesquisadores de *data mining* propuseram diferentes abordagens para gerenciamento desse processo. Os processos básicos propostos por todos eles são relativamente semelhantes. Faremos uma breve análise de alguns dos métodos propostos.

Um desses modelos, o CRISP-DM (Cross-Industry Standard Process for *Data Mining*), foi proposto em meados dos anos 90 por um consórcio de empresas européias para atuar como um modelo de processo padrão não patenteado para *data mining*. A Figura 4.2 ilustra esse processo proposto, que inclui uma sequência de etapas (que não são totalmente sequenciais) para os projetos de *data mining*. Muito trabalho repetido é realizado, particularmente nas primeiras etapas de desenvolvimento de um melhor entendimento das necessidades de negócios e dos dados disponíveis. Não é incomum um projeto de *data mining* tomar até 60% do tempo estimado no desenvolvimento do entendimento de dados e negócios e depois na preparação dos dados para etapas de modelagem reais. Quase sempre, o processo de modelagem leva a uma necessidade de preparação de dados mais detalhada.

Um outro processo que foi proposto para os projetos de *data mining* é baseado na conhecida metodologia Six Sigma. Ela é uma metodologia bem estruturada orientada a dados para eliminação de defeitos, desperdício e problemas de controle de qualidade de todos os tipos na fabricação, entrega de serviço, administração e em outras atividades de negócios. Recentemente, esse modelo tornou-se muito popular devido às suas implementações bem-sucedidas em vários setores nos Estados Unidos e parece estar ganhando preferência mundialmente. No contexto de *data mining*, as cinco etapas do modelo DMAIC (Define [definir], Measure [avaliar], Analyze [analisar], Improve [melhorar] e Control [controlar]) são dadas na Figura 4.3. A idéia básica dessas etapas é ter uma noção preliminar do que esperar, aplicar medição em cada etapa e desenvolver mecanismos de *feedback*. Os processos Six Sigma têm sido usados nos setores de produção em todo o mundo para melhoria da qualidade e controles de processo. Defensores do processo Six Sigma para *data mining* argumentam que o mesmo tipo de eficiência pode ser alcançado em projetos de *data mining*.

Por fim, o SAS Institute, um fornecedor de tecnologias e serviços de *data mining*, propôs um modelo que é um tanto parecido com o Six Sigma. É chamado de SEMMA (*Sample* [selecionar],

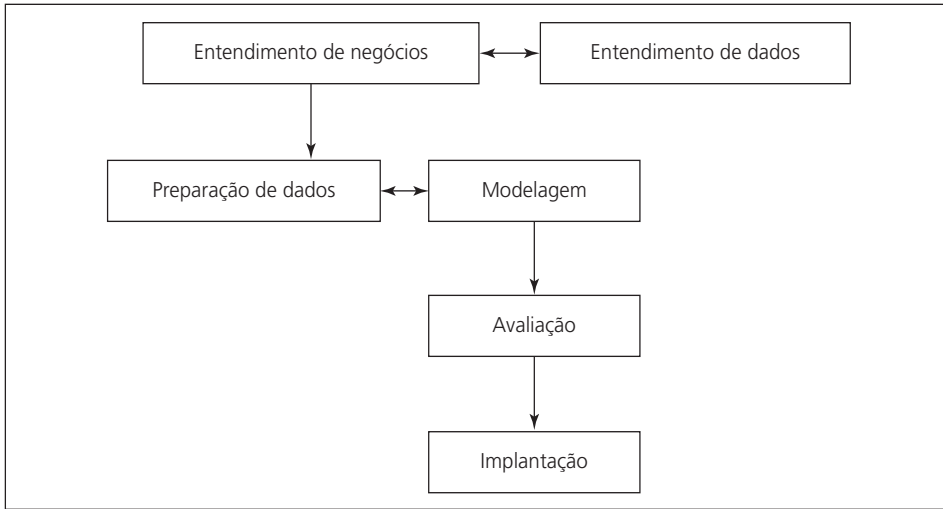


Figura 4.2 Processo de *data mining* recomendado pelo CRISP-DM.



Figura 4.3 Processo de *data mining* baseado no Six Sigma.

Explore [explorar], *Modify* [modificar], *Model* [modelar] e *Assess* [avaliar]). A Figura 4.4 ilustra o processo proposto. O foco do SEMMA é nas atividades técnicas de um projeto de *data mining*.

Como destacado pela StatSoft, “todos esses modelos estão preocupados em como integrar a metodologia de *data mining* em uma empresa, como “converter dados em informação”, como envolver acionistas importantes e como disseminar a informação de forma que possa ser facilmente convertida pelos acionistas em recursos para uma tomada de decisão estratégica”, (statsoft.com 2006). Portanto, o objetivo fundamental dos diferentes processos é muito parecido.

Freqüentemente, alguns adeptos usam o termo *descoberta de conhecimento em bancos de dados* como sinônimo para *data mining*. Fayyad et al. (1996) definiram **descoberta de conhecimento em bancos de dados (KDD)** como um processo que usa os métodos de *data mining* para encontrar informações e padrões úteis nos dados, em oposição ao *data mining*, que envolve o uso de algoritmos para identificar padrões em dados provenientes do processo de KDD. KDD é um processo abrangente que abarca o *data mining*. O material para o processo de KDD consiste de dados organizacionais. Um *data warehouse* empresarial permite que a KDD seja implementada de maneira eficiente, pois fornece uma única fonte para os dados a serem explorados. Dunham (2003) resumiu o processo de KDD como sendo composto pelas seguintes etapas:

1. **Seleção.** Essa etapa envolve a identificação dos dados que serão considerados no processo de *data mining*.



Figura 4.4 Processo SEMMA de *data mining*.

INSIGHTS DE TECNOLOGIA 4.6

Os mitos sobre data mining

O *data mining* é uma ferramenta de análise poderosa que permite aos executivos avançar da descrição do comportamento histórico do cliente para predição do futuro. Ele descobre padrões que revelam mistérios sobre o comportamento do cliente. Os

resultados de *data mining* podem ser usados para aumentar lucros, reduzir despesas, identificar fraude e oportunidades de negócios, oferecendo nova vantagem competitiva. Existem vários mitos sobre essa ferramenta, incluindo:

Mito	Realidade
O <i>data mining</i> fornece predições imediatas como bola de cristal.	O <i>data mining</i> é um processo com várias etapas que exige projeto e uso proativos e calculados.
O <i>data mining</i> ainda não é viável para aplicações de negócios.	A tecnologia atual está pronta para escolher aproximadamente qualquer negócio.
O <i>data mining</i> exige um banco de dados dedicado e distinto.	Devido aos avanços na tecnologia de banco de dados, um banco de dados dedicado não é necessário, embora seja desejável.
Somente aqueles com formação avançada podem fazer o <i>data mining</i> .	Ferramentas baseadas na Web mais recentes permitem que gerentes de todos os níveis educacionais realizem o <i>data mining</i> .
O <i>data mining</i> é apenas para grandes empresas que possuem pilhas de dados de cliente.	Se os dados refletem exatamente o negócio ou seus clientes, uma empresa pode usar o <i>data mining</i> .

Os visionários do *data mining* têm obtido significativa vantagem competitiva ao entender que esses mitos são apenas isso: mitos.

Fontes: adaptado de A. Zaima, “The Five Myths of Data Mining”, *What Works: Best Practices in Business Intelligence and Data Warehousing*, Vol. 15, The Data Warehousing Institute, Chatsworth, CA, June 2003, pp. 42–43 and D. Kuonen, “Data Mining And Statistics: What Is The Connection?” *TDAN.com*, October 2004. tdan.com/i030fe01.htm (acessado em outubro de 2006).

INSIGHTS DE TECNOLOGIA 4.7

Erros crassos em data mining

Muitas vezes, os 10 erros de *data mining* a seguir são cometidos na prática, e deve-se tentar evitá-los:

1. Selecionar o problema errado para *data mining*.
2. Ignorar o que seu patrocinador pensa sobre o que é *data mining* e o que realmente pode e não pode fazer.
3. Não dedicar tempo suficiente para preparação dos dados. Exige mais esforço do que normalmente é deduzido.
4. Olhar apenas resultados reunidos e não os registros individuais. O IMS DB2 da IBM consegue destacar registros individuais de interesse.
5. Ser negligente quanto ao controle de procedimento e resultados do *data mining*.

6. Ignorar descobertas suspeitas e ir adiante rapidamente.
7. Executar algoritmos de mining de modo repetido e cego. É importante pensar bastante sobre o próximo estágio da análise de dados. O *data mining* é uma atividade com muita participação ativa.
8. Acreditar em tudo que se ouve sobre os dados.
9. Acreditar em tudo que se ouve sobre sua própria análise de *data mining*.
10. Medir seus resultados de maneira diferente daquela que seu patrocinador usa para medir os dele.

Fontes: adaptado de D. Skalak, “Data Mining Blunders Exposed!” *DB2 Magazine*, Vol.6, No. 2, Second Quarter 2001, pp. 10–13; and R. Shultz, *Live from NCDM: Tales of Database Buffoonery*, December 7, 2004, directmag.com/news/ncdm-12-07-04/index.html (acessado em abril de 2006).

2. *Pré-processamento*. Dados incorretos e faltantes devem ser tratados; isso envolve correção e/ou uso de valores preditos.
3. *Transformação*. Os dados devem ser convertidos em um único formato comum para processamento; isso pode envolver codificação de dados ou redução no número de variáveis com as quais lidar.
4. *Data mining*. Algoritmos são aplicados aos dados transformados a fim de produzir um resultado.
5. *Interpretação/avaliação*. Para ser útil, os resultados devem ser apresentados de uma maneira que seja expressiva para o usuário.

Como se pode ver a partir da apresentação de vários modelos de processo, o *data mining* é repetitivo porque os data miners aprendem com a experiência. Na verdade, todo o processo de descoberta é repetitivo. Thomas A. Edison ironizou dizendo que falhou 100 vezes ao inventar a lâmpada antes de ter sucesso. Então, parecido com o processo do trabalho do famoso inventor, o *data mining* é um processo de descoberta. É um processo experimental que exige um projeto experimental confiável.

Uma série de mitos sobre *data mining* foi criada. Nós os descrevemos nos *Insights* de tecnologia 4.6. Muitos deles refletem a maneira como o *data mining* é usado na prática. Por exemplo, normalmente os métodos de *data mining* são usados principalmente por equipes de TI, gerentes e consultores/analistas porque é muito difícil para pessoas não especializadas interpretar os resultados. Mas isso está mudando à medida que as ferramentas tornam-se mais fáceis de usar. Veja os *Insights* de tecnologia 4.7 para descobrir erros específicos cometidos pelos data miners na prática, pois muitas vezes eles não entendem o processo embora entendam os resultados esperados.

Desde os ataques terroristas em 11 de setembro de 2001, vários avanços têm ocorrido na utilização de métodos de *data mining* pelos órgãos policiais para rastrear terrorismo e crimes em geral. Entretanto, isso tem levantado muitas questões éticas, legais e de privacidade. Aliás, também tem se tornado uma grande questão política. Ao desenvolver aplicações de *data mining*, não podemos ignorar esses aspectos. Veja o Caso de aplicação 4.7 para detalhes e um exemplo.

CASO DE APLICAÇÃO 4.7

Desdobramentos e avanços do *data mining* do DHS no cumprimento da lei

No fim de 2002, John Poindexter, ex-chefe do Conselho de Segurança Nacional, gerou uma controvérsia com sua proposta para um novo Departamento de Informação (IAO) dentro do Pentágono. Críticas destruíram os planos de Poindexter para fazer *data mining* em vários registros de crédito, bancários e de compras dos cidadãos norte-americanos em nome da detecção de possíveis padrões comportamentais terroristas.

Na realidade, órgãos como o Departamento Nacional de Reconhecimento (NRO) e a Agência de Segurança Nacional (NSA) têm feito isso durante anos, e nos meados de 2003, o Comando do Norte também começou a fazer. Na prática, as ferramentas de *data mining* usadas pelas agências de inteligência nacional já eram usadas pelos órgãos policiais norte-america-

nos. As ferramentas transferiram dados do Comando Espacial Norte-Americano para o Comando do Norte, e de lá para o DHS, mostrando tanto a base da tecnologia comum para todos os ambientes quanto as possíveis preocupações com as liberdades civis inerentes nessas transferências de tecnologia. Uma melhor coordenação do serviço secreto com polícias estaduais e locais, ao mesmo tempo respeitando as liberdades civis, é um assunto importante.

O NRO e a NSA usam ferramentas comerciais de banco de dados em grande escala e ferramentas especializadas para reconhecimento de padrões. Os fornecedores militares são responsáveis pela integração das ferramentas em pacotes de *software* que são úteis para agências de inteligência. Muitos trabalha-

ram com agências pertencentes ao DHS antes de o departamento ser formado, no fim de 2002. Eles implantaram as ferramentas para fiscalização interna de drogas e serviços contraterrorismo através de canais do Comando do Norte e DHS.

Por exemplo, o Sistema de Análise Temporal ativado para Web (WebTAS) da Northrop-Grumman foi desenvolvido em conjunto com os Laboratórios de Pesquisa da Força Aérea e usado durante a operação no Iraque. Ele está disponível para coalizões de inteligência de polícia regional através do DHS. O WebTAS exibe mapas e *links* correspondentes às relações entre alvos. Ao clicar em um *link*, tem-se acesso a bancos de dados relacionados que podem dizer a um analista, por exemplo, todas as chamadas que o alvo fez nos últimos dias. Para obter padrões que podem estar escondidos em ruídos de muita informação, um mecanismo integrado de correlação comportamental prevê possíveis tendências a fim de desenvolver situações e circunstâncias indicadoras que identificam problemas para obtenção de maior inteligência.

Nova cooperação entre os investigadores do DHS, especialmente em *data mining*, está produzindo grandes avanços em casos de não terrorismo, incluindo a morte de 19 estrangeiros ilegais encontrados apertados dentro de um trailer em Victoria, Texas, em 14 de maio de 2003. Detectadas por técnicas de *data mining*, transferências bancárias e ligações telefônicas feitas pelas vítimas e mais de 50 testemunhas que sobreviveram ao incidente levaram as autoridades a um residente legal norte-americano que eles acreditam liderar um grupo de contrabando que traz, por uma grande taxa, estrangeiros pela fronteira entre México e cidades dos EUA. Co-conspiradores em todo o território norte-americano também foram identificados. Após a suspeita fugir do país, foi atraída por uma operação policial secreta para Honduras, onde foi presa e extraditada para os Estados Unidos.

Fontes: adaptado de L. Wirbel, “Data Mining Comes Down to Earth”, *Electronic Engineering Times*, No. 1270, May 19, 2003, pp. 18, 22; and M. Hosenball, “Crime Breakthrough”, *Newsweek*, Vol. 141, No. 26, June 20, 2003, p. 9.

Questões de revisão da Seção 4.4

1. Faça uma breve descrição sobre o modelo CRISP-DM para o processo de *data mining*.
2. Quais são as etapas de *data mining*, de acordo com o método Six Sigma?
3. Quais são as etapas recomendadas pelo modelo SEMMA do SAS Institute para o processo de *data mining*?
4. Defina *KDD*.
5. Identifique alguns dos mitos sobre *data mining*.
6. Quais são alguns dos erros comuns cometidos nos projetos de *data mining*?

4.5

TEXT MINING

Text mining é a aplicação de *data mining* em arquivos de texto não estruturados ou menos estruturados. O *data mining* aproveita-se da infra-estrutura de dados armazenados para extrair informações adicionais úteis. Por exemplo, ao aplicar *data mining* a um banco de dados de clientes, um analista pode descobrir que qualquer pessoa que compra o produto A também compra os produtos B e C seis meses depois. O *text mining* funciona com informações menos estruturadas. Raramente os documentos têm uma estrutura interna sólida, e quando têm, frequentemente está focada no formato do documento ao invés do conteúdo do documento. O *text mining* ajuda as empresas a:

- Encontrar o conteúdo “escondido” dos documentos, incluindo relações adicionais úteis.
- Relacionar documentos ao longo de prévias divisões despercebidas (p.ex., descobrir que os clientes em duas divisões distintas de produto têm as mesmas características).
- Agrupar documentos por assuntos em comum (p.ex., todos os clientes de uma companhia de seguros que têm reclamações semelhantes e cancelaram suas apólices).

O *text mining* não é semelhante a um mecanismo de busca na Web. Em uma busca, estamos tentando encontrar o que os outros prepararam. Com o *text mining*, queremos descobrir novos padrões, pedaços de informação que podem não ser óbvios ou conhecidos.

Basicamente, o *text mining* envolve a geração de índices numéricos significativos a partir de textos não estruturados e, em seguida, processa esses índices, usando os vários algoritmos de *data mining* apresentados nas seções anteriores deste capítulo. Ao criar esses índices numéricos, podemos compactar documentos segundo conceitos-chave, agrupar documentos com base na semelhança, visualizar relações entre documentos, e assim por diante. De fato, empresas agora estão reconhecendo que um grande recurso de vantagem competitiva é o conhecimento não estruturado contido em seu repositório de informações na forma de documentos, memorandos, *e-mails*, políticas e procedimentos, minutas de reuniões, etc. Toda essa informação textual precisa ser codificada e extraída para que as ferramentas preditivas de *data mining* possam ajudar a empresa a gerar valor real desses locais de armazenamento.

Alguns empresários usam *text mining* como um termo geral para indicar todos os tipos de processamento de texto. Mas pelo menos três especialidades apareceram nesse campo: recuperação, extração e resumo da informação. Primeiro, *recuperação de informação* no contexto de *text mining* refere-se a consulta e localização de texto e à apresentação da informação textual. Além disso, o processamento de linguagem natural ou a lingüística computacional está acostumado a analisar e processar textos. Geralmente, isso é referido como *extração de informação*. Programas têm sido desenvolvidos, por exemplo, para ler milhares de currículos e extrair as informações-chave de maneira automática, como nomes, endereços e habilidades. Existem também programas que fornecem um resumo automático dos documentos. Uma demonstração muito boa desse recurso pode ser vista em newsinessence.com. De acordo com o *site*, o sistema NewsinEssence coleta documentos de uma série de *sites*, cria *clusters* baseados em assuntos, e resume cada *cluster*. Sistemas como o NewsinEssence são bons exemplos de extração de informação e ilustram somente um vislumbre da capacidade do processamento de linguagem natural.

- Análise qualitativa de documentos para detectar fraude (veja Cao et al., 2003). Os documentos são decompostos em índices numéricos e, em seguida, várias técnicas de *data mining*, como redes neurais, são usadas para aprender e detectar mentiras em documentos.

Talvez uma das maiores áreas de aplicação para *text mining*, em geral, e extração de informações, em particular, seja a das ciências biológicas. A informação textual inclui coleta e compilação de dados. Isso resultou em muitas atividades de aplicação de técnicas de *data mining* no campo da biologia. Veja Witte (2006) para obter exemplos. Segurança nacional é outra área onde esse tipo de crescimento nas aplicações é evidente. Na próxima seção, discutiremos alguns exemplos nos quais *Web mining* e *text mining* são combinados. O Caso de aplicação 4.8 descreve uma aplicação de *text mining* no setor aéreo.

CASO DE APLICAÇÃO 4.8

Voando pelo texto

O *text mining* provou ser uma ferramenta valiosa na extração de conhecimento organizacional proveniente de relatórios em formato digital. Analistas estão usando *software* de *text mining* para se concentrar em áreas problemáticas essenciais através da identificação de padrões. Empresas do setor aéreo, por exemplo, podem aplicar *text mining* em relatórios de incidente para aumentar a qualidade do conhecimento organizacional. Elas podem investigar problemas

mecânicos, organizacionais e comportamentais em tempo hábil através do uso de *text mining*.

Companhias aéreas trabalham com uma análise eficiente e sistemática das operações. Um relatório de incidente é preparado toda vez que ocorrer um evento que resulte em um problema. Usando o *text mining*, questões-chave podem ser identificadas a partir de um grande número de relatórios de incidente. Os enormes bancos de dados que as companhias aéreas

possuem têm uma interpretação humana limitada, e a terminologia aparenta ser diferente para um computador do que para um humano.

Por exemplo, os dados da Aer Lingus (**aerlingus.com**) entre janeiro de 1998 e dezembro de 2003 foram usados para encontrar padrões e correlações, levando a uma análise adicional e ao desenvolvimento de um modelo. A Aer Lingus usou o *software* de data e text mining PolyAnalyst da Megaputer (**megaputer.com**). Seu objetivo era visualizar um processo que os investigadores pudessem usar regularmente para identificar padrões e associações em tipos de incidentes, locais, horários e outros detalhes sobre os incidentes. Os termos mais frequentes foram identificados inicialmente. O PolyAnalyst possui um dicionário de termos que não está completo, mas fornece um ponto de partida valioso para análise de texto. Ele pode gerar uma lista de termos-chave (ou seus equivalentes semânticos) que ocorrem nos dados. Um relatório chamado *relatório de termos frequentes* é criado e contém os termos identificados com suas frequências. O objetivo é identificar *clusters* interes-

santes. Um resumo descritivo traz um conjunto de termos que dividem as descrições em grupos expressivos. O termo-chave *derramamento*, por exemplo, pode ser associado a quatro outros termos-chave: *comida*, *combustível*, *produto químico* e *banheiro*. A partir dos termos-chave relacionados a derramamento, *comida* é semanticamente associado a *café*, *chá* e *bebida alcoólica*. Portanto, *comida* torna-se a categoria nó, e os diferentes produtos alimentícios relatados como derramados são combinados com comida.

Text mining de relatórios de incidentes aéreos podem identificar incidentes que podem resultar em transtorno. O *text mining* pode ser usado com um grande conjunto de relatórios de dados sobre incidentes a fim de validar teorias predefinidas e escolher novos padrões de conhecimento.

Fontes: J. Froelich, S. Ananyan, and D.L. Olson, “Business Intelligence Through Text Mining”, *Business Intelligence Journal*, Vol.10, No. 1, Winter 2005, p. 43–50; and *Gain Full Value from Text Responses*, spss.com/textanalysis_surveys/ (acessado em abril de 2006).

Como explorar textos

Ellingsworth e Sullivan (2003) descreveram o processo de *text mining*. Eles também descreveram como a companhia de seguros Fireman's Fund usa o *text mining* para ajudar a prever ações esperadas e compreender por que os resultados divergem das previsões. O *text mining* é usado para extrair entidades e objetos para análise de frequência, identificar arquivos que tem determinados atributos para análise estatística mais detalhada e criar integralmente novos itens de dados para modelagem preditiva. O primeiro desses três métodos foi usado para enfrentar os processos judiciais envolvendo os pneus Firestone nos utilitários da Ford. No Arquivo online W4.1, descrevemos os detalhes de uma aplicação efetiva de *text mining* na área farmacêutica.

Extração de termo é a forma mais básica de *text mining*. Como todas as outras técnicas de *data mining*, ela mapeia informações de dados não estruturados em um formato estruturado. A estrutura de dados mais simples em *text mining* é o vetor de características, ou lista de palavras ponderada. As palavras mais importantes em um texto são listadas, junto com um indicador de sua importância relativa. O texto é reduzido a uma lista de termos e pesos. Talvez toda a semântica do texto não seja apresentada, mas os conceitos-chave são identificados. Para isso, *text mining* envolve as seguintes etapas:

1. *Eliminar palavras comumente usadas (o, a, e, outro[a]).* Normalmente são chamadas de *stopwords*.
2. *Substituir as palavras por seus radicais (p.ex., eliminar plural e várias conjugações e declinações).* Nessa etapa, os termos *telefonado*, *telefonando* e *telefona* seriam mapeados por *telefon*. Eles são chamados de *algoritmos de stemming*.
3. *Considerar sinônimos e frases.* Palavras que são sinônimas precisam ser combinadas de alguma maneira. Por exemplo, *estudante* e *aluno* precisam ser colocados juntos. Também é importante levar em consideração frases. Como destacado pela StatSoft, *Microsoft Windows* refere-se ao sistema operacional do computador, mas sozinha, a palavra *Windows* (janela) pode ser mais relevante para o projeto de uma casa (statsoft.com 2006).

4. *Calcular a relevância dos termos restantes.* O método mais comum é calcular a frequência com a qual as palavras aparecem. Existem duas medidas usuais: a frequência do termo, ou *fator tf*, mede o número real de vezes que uma palavra aparece em um documento, e a frequência inversa do documento, ou *fator idf*, indica o número de vezes que a palavra aparece em todos os documentos de um conjunto. O raciocínio é que um fator *tf* alto eleva o peso, enquanto que um fator *idf* alto diminui o peso porque os termos que ocorrem com frequência em todos os documentos seriam palavras comuns para o setor e não seriam consideradas importantes.

Suponha, por exemplo, que a análise de um parágrafo leve à observação de que existem uns 20 termos com 28 ocorrências quando consideramos palavras comuns. Segue uma lista de termos que aparecem mais de uma vez, junto com suas frequências relativas (fatores *tf*) em um total de 28:

<i>Termo</i>	<i>Fator do termo</i>
dados	0,0714
estrutura	0,0714
termo	0,0714
texto	0,0714
<i>text mining</i>	0,1429
peso	0,0714

Quando se considera todas as palavras importantes do parágrafo, elas abrangem metade da sua importância total e poderiam ser usadas para identificar sua semântica. Claramente, o parágrafo é sobre *text mining* (peso = 0,1479) e envolve texto e dados com estrutura e peso.

Existem muitos componentes em um sistema de *text mining*, incluindo:

- Um sistema para manuseio de documentos em várias formas (p.ex., texto puro, formatos de processador de texto, PDF) provenientes de diferentes fontes (p.ex., arquivos, formato Web, *e-mails*).
- Componentes usados para processar esses documentos e criar arquivos de dados que possam, então, ser explorados. Incluem “segmentadores sentenciais, classe gramatical das palavras, etiquetador,..., parser completo,... resumidores,” etc (Witte, 2006, p. 134-138).
- Ferramentas de *data mining* como algoritmos de *cluster*, algoritmos classificadores, etc.

Existem dois modelos gerais com código aberto disponíveis para as duas primeiras tarefas. Um modelo popular é chamado de GATE (General Architecture for Text Engineering) e está disponível no site da Universidade de Sheffield (gate.ac.uk), Reino Unido. Outra plataforma, suportada pela IBM, é chamada de UIMA (Unstructured Information Management Architecture) e está disponível no site research.ibm.com/UIMA/. Além dessas ferramentas, muitos fornecedores de *data mining* oferecem recursos de *text mining* nos seus pacotes de *software*. Como essa área ainda está sujeita a mais pesquisa e desenvolvimento, os recursos do *software* mudam rapidamente. Segue uma lista de algumas ferramentas de *text mining* populares e seus fornecedores:

- SAS Text Miner (sas.com)
- IBM Intelligent Miner for Text (ibm.com)
- SPSSLexiquest (spss.com)
- Insightful Miner for Text (insightful.com)
- Megaputer Intelligence TextAnalyst (megaputer.com)
- StatSoft Data Miner (statsoft.com)

O Arquivo *online* W4.2 fornece instruções passo a passo para um projeto de *text mining* completo usando os recursos do Text Miner e do Data Miner da StatSoft. Esse arquivo mostra como analisar

os documentos. O exemplo específico está relacionado à análise das revisões de carro preparadas por pessoas muito diferentes. Esse tipo de “*opinion mining*” é outra aplicação usual de *text mining*.

Dado que possivelmente 80% de todos os dados não numéricos que coletamos e armazenamos estão em forma de texto, é natural que o *text mining* esteja surgindo como uma grande área de crescimento. Embora ainda não tenhamos recursos plenos para processamento de linguagem natural, já foi feito muito progresso nessa área nos últimos anos. É uma área onde existe potencial significativo para a próxima geração de aplicações úteis.

Questões de revisão da Seção 4.5

1. O que é *text mining*?
2. Quais são algumas das especialidades de processamento de texto relacionadas ao *text mining*?
3. Dê exemplos de aplicações de extração automática de informação.
4. Identifique algumas aplicações atuais e potenciais de *text mining*.
5. Quais são algumas das etapas básicas para realização de *text mining*?
6. Identifique os componentes básicos de qualquer sistema de *text mining*.

4.6

WEB MINING

Provavelmente a Web seja o maior repositório de dados/textos do mundo, e a quantidade de informação armazenada cresce a cada dia. Existe informação interessante na Web sobre páginas vinculadas a outras páginas, sobre como inúmeras pessoas têm *hyperlinks* com uma página da Web específica em seus próprios *sites*, e como um site privado é organizado. Além disso, cada visitante de um *website*, cada pesquisa em um mecanismo de busca, cada clique em um *link* e cada operação em um site de comércio eletrônico leva à criação de dados adicionais. A análise dessa informação pode nos ajudar a fazer um melhor uso dos *websites* e também nos ajudar a fornecer relação e valor melhores para os visitantes de nossos *sites*.

A **Web mining** pode ser definida como a descoberta e análise de informações úteis e interessantes provenientes da Web, sobre a Web e, geralmente, através de ferramentas baseadas na Web. O termo *Web mining* foi usado pela primeira vez por Etzioni (1996); atualmente, muitas associações, revistas e muitos livros focam-se em *data mining* da Web. Ela representa uma das áreas em constante desenvolvimento da tecnologia e das práticas de negócios. A Figura 4.5 apresenta três diferentes áreas de *Web mining*: de conteúdo, de estrutura e de uso.

Web mining de conteúdo refere-se à extração de informações úteis de páginas Web. Os documentos podem ser extraídos em algum formato legível por máquina para que técnicas automá-

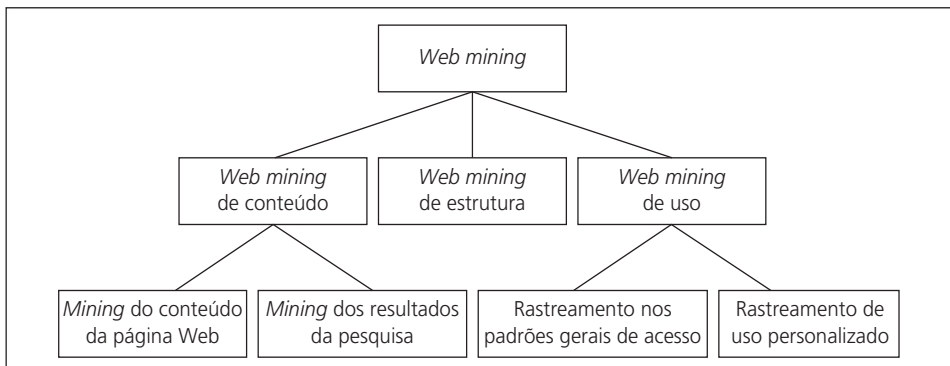


Figura 4.5 Tipos de *Web mining*.

Fonte: adaptado de P. Galeas, *Web Mining*, galeas.de/webmining.htm (acessado em abril de 2006).

ticas possam gerar alguma informação sobre as páginas. Os robos de pesquisa na Web são usados para ler todo o conteúdo de um *website* de forma automática. A informação automatizada pode incluir características de documento semelhantes às usadas em *text mining*, mas pode incluir conceitos adicionais, como hierarquia de documento. *Web mining* de conteúdo também pode ser usada para melhorar os resultados produzidos pelos mecanismos de busca. Por exemplo, Turetken e Sharda (2004) descreveram um sistema de visualização que pega os resultados de uma pesquisa, a partir de um mecanismo de busca como o Google, lê os 100 principais documentos, agrupa esses documentos ao processá-los pelo Intelligent Text Miner da IBM e, então, apresenta os resultados em um formato gráfico baseado em visão olho-de-peixe.

Web mining de estrutura refere-se ao desenvolvimento de informações úteis a partir de *links* incluídos nos documentos Web. Essa informação pode ser útil na determinação da popularidade de um documento, um pilar do algoritmo de classificação de página que é central para mecanismos de busca como o Google. Da mesma forma que os *links* que levam a um documento podem indicar a popularidade do documento, *links* dentro do documento podem indicar a profundidade de cobertura de um assunto. Kleinberg (1998) identificou um mecanismo para encontrar “*hubs*” e “*autoridades*” através da análise de *hyperlink*. *Hubs* são páginas que apontam para muitas autoridades no campo. Por outro lado, *páginas de autoridade* são páginas que estão vinculadas a muitos *hubs*. Um bom entendimento da estrutura de um *website* é útil na determinação de quais páginas devem ser adicionadas a um acervo de *sites*. Uma duplicação significativa de *links* Web em uma página pode sugerir que não é necessário vincular dois *sites* diferentes, por exemplo. A análise de *links* é muito importante para a compreensão das inter-relações. Algumas vezes, pode até levar ao desenvolvimento de um melhor entendimento de uma comunidade, clã ou grupo Web específico. Zhou et al. (2005) descreveram um grande projeto no qual eles usaram tanto *Web mining* de conteúdo quanto *Web mining* de estrutura para melhor entender como os grupos extremistas norte-americanos estão relacionados (veja o Caso de aplicação 4.9).

CASO DE APLICAÇÃO 4.9

Apanhado pela rede

Normalmente procuramos as respostas para nossos problemas fora do nosso meio. Mas, em muitos casos, o aborrecimento pode originar-se de dentro. Ao agir contra o terrorismo global, grupos extremistas nacionais passam despercebidos. Entretanto, extremistas internos com ideologias diferentes representam uma ameaça significativa à segurança devido à informação que possuem e também à capacidade crescente, através do uso da Internet, de serem ouvidos por outros grupos extremistas internacionais.

Controlar o conteúdo disponível na Internet é difícil. Pesquisadores e autoridades precisam de ferramentas de ponta para analisar e monitorar as atividades de grupos extremistas. Pesquisadores da Universidade do Arizona, com apoio do DHS e de outros órgãos, desenvolveram uma metodologia de *Web mining* para encontrar e analisar *websites* operados por esses extremistas internos a fim de aumentar o entendimento sobre os grupos através do uso da In-

ternet. Grupos extremistas usam a Internet para se comunicar, acessar mensagens particulares e arrecadar dinheiro *online*. Inicialmente, a metodologia de pesquisa reúne um acervo de alta qualidade sobre *websites* de extremistas e terroristas relevantes. A análise de *hyperlink* é realizada, o que leva a outros *websites* de terroristas e extremistas. A interligação com outros *sites* é fundamental para avaliar a semelhança nos objetivos de vários grupos. A próxima etapa na metodologia é a análise de conteúdo, que posteriormente codifica esses *websites*, com base em diversos atributos, como comunicações, arrecadação de recursos e compartilhamento de ideologia, só pra citar alguns.

Com base nas análises de *link* e conteúdo, os pesquisadores se restringiram a 97 *sites* de grupos extremistas e racistas norte-americanos. Frequentemente, os vínculos entre comunidades não representam necessariamente qualquer cooperação entre elas. Entretanto, encontrar muitos vínculos entre grupos com

interesses em comum ajuda a agrupar as comunidades sob uma mesma bandeira. O principal objetivo de grupos extremistas e racistas é trazer à tona suas definições de vida e objetivos. Pesquisas adicionais usando *data mining* para automatizar o processo têm um propósito global, ao tentar estabelecer vínculos entre grupos extremistas e racistas internacionais e norte-americanos.

Fontes: Y. Zhou, E. Reid, J. Qin, H. Chen, and G. Lai, “U.S. Domestic Extremist Groups on the Web: Link and Content Analysis”, *IEEE Intelligent Systems*, Vol. 20, No. 5, September/October 2005, pp. 44–51, ai.arizona.edu/go/intranet/papers/Colorprint.pdf (acessado em abril de 2006); and P.B. Gerstenfeld, D.R. Grant, and C.-P. Chiang, “Hate Online: A Content Analysis of Extremist Internet Sites”, *Analyses of Social Issues and Public Policy*, Vol. 3, No. 1, December 2003, pp. 29–44.

A **Web mining de uso** refere-se à extração de informações úteis a partir de dados gerados por visitas e operações, entre outros, em páginas Web. Srivastava et al. (2002) afirmaram que pelo menos três tipos de dados são gerados por visitas a páginas Web:

- Dados gerados automaticamente que são armazenados em registros de *acesso* ao servidor, registros de *referência*, registros de *agente* e *cookies* do lado cliente.
- Perfis do usuário
- Metadados como atributos de página, atributos de conteúdo e dados de uso

Analisar as informações coletadas pelos servidores Web pode nos ajudar a entender o comportamento do usuário. Isso geralmente é chamado de *análise de acesso na Web*. Através do uso de técnicas de *data mining* abordadas anteriormente neste capítulo, uma empresa pode ser capaz de discernir padrões interessantes a partir do fluxo de cliques. Por exemplo, ela pode saber que 60% dos visitantes que procuraram “hotéis em Maui” anteriormente procuraram “passagens aéreas para Maui”. Essa informação pode ser muito útil para decidir onde colocar as propagandas *online*. Análise de acesso na Web também pode ser útil para saber *quando* os visitantes acessam o *site*. Por exemplo, se você soubesse que 70% dos *downloads* de *software* no seu *site* aconteceram entre 7 e 11 horas no horário leste, poderia planejar suporte ao cliente e largura de banda da rede melhores durante esses horários. Srivastava (2006) listou os seguintes exemplos de uso para a *Web mining*:

- Determinar o valor do tempo de vida dos clientes
- Definir estratégias de *marketing* cruzado para produtos
- Avaliar campanhas promocionais
- Direcionar anúncios e cupons eletrônicos para grupos de usuários, com base nos seus padrões de acesso
- Prever o comportamento do usuário com base em regras previamente aprendidas e perfis de usuários
- Apresentar informações dinâmicas para os usuários com base nos seus interesses e perfis

A Amazon.com fornece um bom exemplo de como um histórico de uso da Web pode ser potencializado dinamicamente. Um usuário registrado que visita novamente a Amazon.com é cumprimentado pelo nome. Essa é uma tarefa simples que envolve o reconhecimento do usuário através da leitura de um *cookie* (ou seja, um pequeno arquivo de texto escrito pelo *website* no computador do visitante). A Amazon.com também apresenta ao usuário uma seleção de produtos em uma loja personalizada, com base nas compras anteriores e em uma análise de associação de usuários semelhantes. Também faz ofertas especiais de “Gold Box” que são boas, mas por tempo limitado. Todas essas sugestões envolvem uma análise detalhada do visitante e também do grupo de iguais do usuário, usando agrupamento, descoberta de sequência de padrões, associação e outras técnicas de *data mining*. A Figura 4.6 ilustra essa personalização na Amazon.com.

Os *software* para análise de uso da Web estão amplamente disponíveis. Além disso, a maioria dos produtos de *data mining* inclui a habilidade de analisar os dados gerados pelos registros de acesso na Web de um *site*. A Tabela 4.5 inclui uma lista parcial de produtos de *software* que são dire-

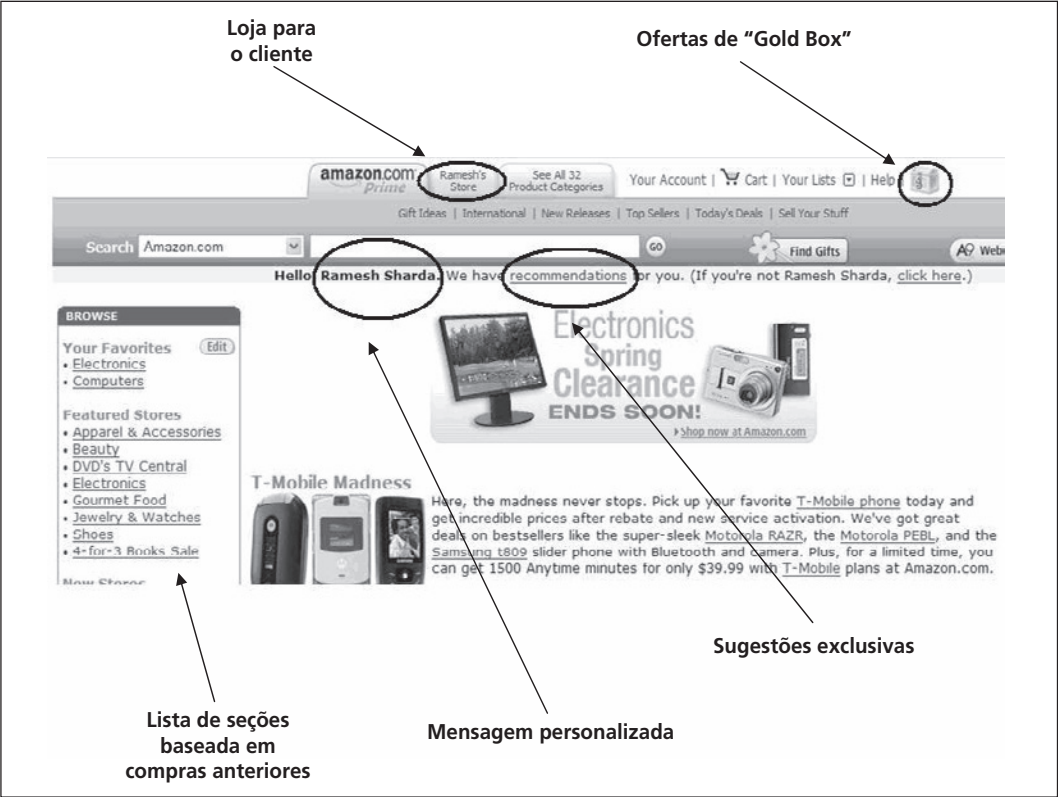


Figura 4.6 Exemplo de personalização usando *Web mining* de uso.

Tabela 4.5 Uma amostra de *softwares* de *Web mining* de uso.

Nome do produto	URL	Descrição
Angoss Knowledge WebMiner	angoss.com	Associa ANGOSS KnowledgeSTUDIO e análise de acesso na Web
ClickTracks	clicktracks.com	Padrões do visitante podem ser mostrados no site
LiveStats da DeepMetrix	deepmetrix.com	Análise de registro em tempo real, demo ao vivo no site
Megaputer WebAnalyst	http://megaputer.com/products/wm.php3	Recursos de <i>data</i> e <i>text mining</i>
MicroStrategy Web Traffic Analysis Module	microstrategy.com/Solutions/Applications/WTAM	Destaque em sinalizadores, análise de conteúdo, e relatórios de análise do visitante da Web
SAS Web Analytics	sas.com/solutions/webanalytics/	Analisa o tráfego do website
SPSS <i>Web Mining</i> for Clementine	spss.com/web_mining_for_clementine	Extração de eventos da Web
WebTrends	webtrends.com	Data mining da informação de tráfego da Web
XML Miner	scientio.com	Um sistema e biblioteca de classe para <i>data mining</i> , e texto expresso em XML, usando regras de sistema especialista em lógica fuzzy

tamente voltados para análise de dados do registro da Web. Mais produtos nessa categoria podem ser encontrados *online* em kdnuggets.com/software/web-mining.html e em outros *sites* semelhantes.

Questões de revisão da Seção 4.6

1. Defina *Web mining*.
2. Identifique as três principais áreas de *Web mining*.
3. Quais são os usos da *Web mining* de estrutura?
4. Quais tipos de informação os servidores Web coletam quando se visita uma página Web e se faz uma atividade interativa?
5. Liste cinco exemplos de observações que um analista de *Web mining* poderia fazer com os dados de uso da Web.
6. Identifique aplicações desenvolvidas pelos maiores *sites* de comércio eletrônico e que podem estar baseadas em *Web mining* de uso.

Destaques do capítulo

- O *data mining* é a descoberta de conhecimento em bancos de dados. Muitas vezes é realizado em dados em *data warehouses*.
- Empresas usam o *data mining* para melhor entender seus clientes e otimizar suas operações.
- Na prática, aplicações de *data mining* estão em todas as áreas de negócios e governamentais, incluindo saúde, finanças, *marketing* e segurança nacional.
- Quatro categorias gerais de algoritmos são usadas em *data mining*: classificação, agrupamento, associação e descoberta de sequência.
- Os métodos de classificação aprendem de exemplos anteriores que contêm entradas e a classe resultante, e são capazes de classificar padrões futuros.
- O agrupamento divide registros de padrão em segmentos ou grupos (*clusters*). Cada membro do segmento compartilha qualidades semelhantes.
- O *data mining* pode ser baseado em hipótese ou baseado em descoberta. O *data mining* baseado em hipótese começa com uma proposição pelo usuário. O *data mining* baseado em descoberta é uma expedição mais ilimitada.
- Uma série de algoritmos diferentes é comumente usada para classificação. Implementações comerciais incluem ID3, C4.5, C5, CART e SPRLINT.
- O uso de árvores de decisão envolve a criação de uma árvore para dividir os dados através da ramificação em diferentes atributos, de modo que o nó folha tenha todos os padrões de uma classe.
- Duas maneiras de determinar as escolhas de ramificação em uma árvore de decisão é usar o índice Gini e o ganho de informação (entropia).
- O índice Gini mede a pureza de uma amostra. Se tudo em uma amostra pertence a uma classe, então o valor do índice Gini é zero.
- Os algoritmos de *cluster* são usados onde os registros de dados não possuem identificadores de classe predefinidos. Isto é, não sabemos a qual classe pertence determinado registro.
- Os algoritmos de *cluster* calculam as medidas de semelhança a fim de agrupar padrões similares mais estreitamente.
- De forma semelhante a outras iniciativas, um projeto de *data mining* deve seguir um processo de gerenciamento de projeto.
- KDD é um processo de uso de *data mining* para descobrir padrões em dados.
- O *text mining* é a aplicação de *data mining* em arquivos de texto. Dado que uma grande quantidade de informação está em forma de texto, *text mining* está crescendo rapidamente.
- O *text mining* está estreitamente relacionado a extração de informação, processamento de linguagem natural e resumo de documentos.
- O *text mining* envolve a criação de índices numéricos a partir de dados não estruturados e, então, a aplicação dos algoritmos de *data mining* a esses índices.
- A *Web mining* pode ser definida como descoberta e análise de informações úteis e interessantes a partir da Web, sobre a Web e, em geral, usando ferramentas baseadas na Web.
- A *Web mining* pode ser vista como um composto de três áreas: *Web mining* de conteúdo, *Web mining* de estrutura e *Web mining* de uso.
- A *Web mining* de conteúdo refere-se à extração automática de informações úteis a partir de páginas Web. Pode ser usada para melhorar os resultados de pesquisa produzidos por mecanismos de busca.
- A *Web mining* de estrutura refere-se à geração de informações interessantes a partir de *links* incluídos em

páginas Web. É usada no algoritmo de classificação de página do Google para, por exemplo, organizar a exibição das páginas.

- A *Web mining* de estrutura também pode ser usada para identificar os membros de uma comunidade específica, e talvez até as funções dos membros na comunidade.
- A *Web mining* de uso refere-se ao desenvolvimento de informações úteis através da análise de registros

do servidor Web, perfis de usuários e informações sobre operações.

- A *Web mining* de uso pode auxiliar em um melhor CRM, personalização, modificações na navegação do site e modelos de negócios aprimorados.
- *Data, text* e *Web mining* surgem como componentes fundamentais das ferramentas de TI da próxima geração para possibilitar às empresas competir com sucesso.

Termos-chave

- agrupamento
- árvore de decisão
- associação
- classificação
- *data mining*
- *data mining* baseado em descoberta
- *data mining* baseado em hipótese
- descoberta de conhecimento em bancos de dados (KDD)
- descoberta de sequência
- entropia
- índice Gini
- *text mining*
- *Web mining* de conteúdo
- *Web mining*
- *Web mining* de estrutura
- *Web mining* de uso

QUESTÕES PARA DISCUSSÃO

1. Explique a relação entre OLAP e *data mining*.
2. Discuta o que uma empresa deve levar em consideração antes de tomar a decisão de comprar um *software* de *data mining*.
3. Distinga *data mining* de outras ferramentas de análise.
4. Defina *data mining* e liste suas principais tecnologias.
5. Diferencie *data mining*, *text mining* e *Web mining*.
6. Distinga KDD de *data mining*. Explique como a Web está afetando os métodos e as tecnologias analíticas de negócios/BI.
7. Discuta o que uma empresa deve levar em consideração antes de tomar a decisão de comprar um *software* de *data mining*.
8. Distinga o *data mining* de outras ferramentas de análise.
9. Explique o processo de *text mining*.
10. Descreva por que a visualização é tão importante em BI.
11. Em quais situações os *software* de classificação e agrupamento devem ser usados juntos?
12. Discuta por que o *text mining* não é a mesma coisa que processamento de linguagem natural.

Exercícios

Exercícios da Teradata University e outros exercícios interativos

1. Acesse teradatastudentnetwork.com. Identifique casos sobre *data mining*. Descreva desenvolvimentos recentes no campo.
2. Acesse teradatastudentnetwork.com ou a URL fornecida pelo seu instrutor. Encontre o seminário na Web sobre *data mining*. Encontre, em especial, um seminário dado por C. Imhoff e T. Zouques. Assista ao seminário na Web. Em seguida, responda as seguintes questões:
 - a. Quais são algumas aplicações interessantes de *data mining*?

- b. Que tipos de benefícios e custos as empresas podem esperar das iniciativas de *data mining*?

3. Para esse exercício, seu objetivo é criar um modelo para identificar entradas ou indicadores que diferenciem clientes de risco dos outros (com base em padrões relacionados a clientes anteriores). Em seguida, use essas entradas para prever novos clientes de risco. Esse caso de exemplo é típico para esse campo.

Os dados amostrais a serem usados nesse exercício estão no Arquivo *online* W6.3, no arquivo CreditRisk.xls. O conjunto de dados tem 45 casos e 15 variáveis relacionados a clientes antigos e atuais que, por várias razões, fizeram um empréstimo no

banco. O conjunto de dados contém diversas informações ligadas a clientes, situação financeira, razão para o empréstimo, emprego, informação demográfica, e a variável resultante ou dependente para capacidade creditícia, classificando cada caso como bom ou ruim, com base em experiência anterior da instituição. Tome 400 casos como casos de treinamento e separe outros 25 para teste. Então, crie um modelo de árvore de decisão para aprender as características do problema. Teste o desempenho do modelo nos outros 25 casos. Registre o desempenho de aprendizado e teste do seu modelo. Prepare um relatório que identifique o modelo de árvore de decisão e parâmetros de treinamento, assim como o desempenho resultante no conjunto de teste. Use qualquer *software* de árvore de decisão.

(Esse exercício é cortesia da StatSoft Inc., baseado em um conjunto de dados em língua alemã proveniente de <ftp://ftp.ics.uci.edu/pub/machine-learning-databases/statlog/german> renomeado CreditRisk e modificado).

Tarefas em grupo e interpretação de papéis

1. Verifique como os novos dispositivos de captura de dados, como as etiquetas de identificação por radiofrequência (RFID), ajudam as empresas a identificar e segmentar com precisão seus clientes para atividades como *marketing* direcionado. Muitas dessas aplicações envolvem *data mining*. Pesquise na bibliografia e na Web e, então, proponha cinco possíveis novas aplicações de *data mining* para a tecnologia de RFID. Quais questões poderiam surgir se a legislação de um país exigisse que tais dispositivos fossem integrados ao corpo de cada pessoa para um sistema de identificação nacional?
2. Entreviste administradores na sua faculdade ou executivos na sua empresa para determinar como as ferramentas de *data warehousing*, *data mining*, OLAP, visualização e BI/DSS poderiam auxiliá-los no trabalho. Escreva uma proposta descrevendo suas descobertas. Inclua estimativas de custo e benefícios no seu relatório.
3. Existe um bom repositório de dados que foi usado para testar o desempenho de muitos algoritmos de aprendizado de máquina. Esse repositório pode ser acessado em ics.uci.edu/~mllearn/MLRepository.html. Alguns conjuntos de dados são destinados para teste dos limites de algoritmos de aprendizado de máquina e comparam seu desempenho com novas abordagens de aprendizado. Entretanto, alguns conjuntos de dados menores podem ser úteis na exploração da funcionalidade de qualquer *software* de *data mining* ou do *software* disponível como complemento deste livro, como o STATISTICA Data-Miner. Faça o *download* de pelo menos um conjunto de dados desse repositório (p.ex., Bancos de dados de verificação de crédito, Banco de dados imobiliário). Em seguida, aplique métodos de árvore de decisão ou agrupamento, conforme necessário. Prepare um relatório com seus resultados. (Alguns desses exercícios podem ser propostos como projetos semestrais para um trabalho final, por exemplo.)
4. Considere o seguinte conjunto de dados, que inclui três atributos e uma classificação para decisões de admissão em um programa de MBA:

GMAT	Média das notas	Pontuação quantitativa do GMAT (percentil)	Decisão
650	2,75	35	Não
580	3,50	70	Não
600	3,50	75	Sim
450	2,95	80	Não
700	3,25	90	Sim
590	3,50	80	Sim
400	3,85	45	Não
640	3,50	75	Sim
540	3,00	60	?
690	2,85	80	?
490	4,00	65	?

- a. Usando os dados mostrados acima como exemplos, desenvolva seu próprio manual prático de regras para tomada de decisão.
- b. Crie uma árvore de decisão usando SPRINT (ou seja, o índice Gini). Você pode criá-la usando cálculos manuais ou uma planilha para realizar os cálculos básicos.
- c. Crie outra árvore de decisão, agora usando a abordagem de entropia e ganho de informação (ID3). Você pode usar uma planilha como calculadora nesse exercício.
- d. Use um programa de *software* para árvore de decisão automática (p.ex., See5) para criar uma árvore para os mesmos dados.
- e. Registre as previsões das últimas três observações de cada abordagem de classificação.
- f. Comente a semelhança e a classificação das abordagens. O que você aprendeu nesse exercício?

Exercícios na Internet

1. Acesse cs.ualberta.ca/~aixplore/, que é um projeto chamado de AI Exploratorium. Clique no *link*

Árvore de decisão (Decision Tree). Leia a narrativa sobre estatística em jogos de basquete. Verifique os dados e, em seguida, crie uma árvore de decisão. Registre suas impressões sobre a precisão dessa árvore de decisão. Além disso, explore os resultados de algoritmos diferentes.

2. Pesquise sobre algumas ferramentas de *data mining* e alguns fornecedores. Comece com **fairisaac.com** e **egain.com**. Consulte também **dmreview.com** e identifique alguns produtos de *data mining* e fornecedores de serviços que não foram mencionados neste capítulo.
3. Encontre casos recentes de aplicações de *data mining* bem-sucedidas. Tente os fornecedores de *data mining* e procure casos ou histórias de sucesso. Prepare um relatório com o resumo de cinco novos estudos de caso.
4. Acesso os sites (em especial SAS, SPSS, Cognos, Teradata, StatSoft e Fair Isaac) e procure em histórias de sucesso por ferramentas de BI (OLAP e *data mining*). O que você encontrou em comum entre as várias histórias de sucesso? Como elas se diferenciam?

Dos Exercícios 5-11, talvez você precise procurar informações em categorias como Estudos de caso, Histórias de sucesso, Estudo do cliente etc. Os *sites* mudam sua organização freqüentemente, por isso concentre-se no objetivo de localizar os relatórios de aplicação nos sites das empresas.

5. Acesse **statsoft.com**. Selecione *Downloads* e faça o *download* de pelos menos três documentos informativos sobre aplicações. Qual dessas aplicações pode ter usado as técnicas de *data/text/Web mining* discutidas neste capítulo?

6. Acesse **sas.com**. Selecione *Downloads* e faça o *download* de pelos menos três documentos informativos sobre aplicações. Qual dessas aplicações pode ter usado as técnicas de *data/text/Web mining* discutidas neste capítulo?

7. Acesse **spss.com**. Selecione *Downloads* e faça o *download* de pelos menos três documentos informativos sobre aplicações. Qual dessas aplicações pode ter usado as técnicas de *data/text/Web mining* discutidas neste capítulo?

8. Acesse **teradata.com**. Selecione *Downloads* e faça o *download* de pelos menos três documentos informativos sobre aplicações. Qual dessas aplicações pode ter usado as técnicas de *data/text/Web mining* discutidas neste capítulo?

9. Acesse **fairisaac.com**. Selecione *Downloads* e faça o *download* de pelos menos três documentos informativos sobre aplicações. Qual dessas aplicações pode ter usado as técnicas de *data/text/Web mining* discutidas neste capítulo?

10. Acesse **salfordsystems.com**. Selecione *Downloads* e faça o *download* de pelos menos três documentos informativos sobre aplicações. Qual dessas aplicações pode ter usado as técnicas de *data/text/Web mining* discutidas neste capítulo?

11. Acesse **rulequest.com**. Selecione *Downloads* e faça o *download* de pelos menos três documentos informativos sobre aplicações. Qual dessas aplicações pode ter usado as técnicas de *data/text/Web mining* discutidas neste capítulo?

12. Acesse **kdnuggets.com**. Explore as seções sobre aplicações e *software*. Encontre nomes de pelo menos três pacotes adicionais para *data mining* e *text mining*.

CASO DE APLICAÇÃO DE FINAL DO CAPÍTULO

Hewlett-Packard e *text mining*

A Hewlett-Packard Company (HP), fundada em 1939 por William R. Hewlett e David Packard, tem sua sede em Palo Alto, Califórnia. A empresa fornece produtos, tecnologias, soluções e serviços para pessoas, pequenas e médias empresas, e grandes empresas com base mundial. A HP também oferece soluções de *software* de gerenciamento que permitem aos clientes empresariais administrar sua infra-estrutura de TI, operações, aplicações, serviços de TI, processos de negócios e também plataformas de grande porte para diversas aplicações. Algumas das categorias conhecidas de produto da HP são computadores comerciais e

pessoais, estações de trabalho, dispositivos portáteis, impressoras jato de tinta, impressoras a *laser*, sistemas digitais de entretenimento, calculadoras e acessórios relacionados, *software* e serviços, fotografia e entretenimento digitais, recursos gráficos, suprimentos de imagem e impressão para *hardware* de impressora, suprimentos de impressão, *scanners*, e produtos para infra-estrutura de rede, incluindo *switches* Ethernet. Os revendedores formam o canal de distribuição da empresa. A HP também vende através de parceiros de distribuição, fabricantes de equipamento original e integradores de sistemas.

TEXT MINING

Os clientes da HP se comunicam com a empresa através de milhares de *e-mails*. A análise de dados estruturados é eficaz na descoberta de parâmetros como, por exemplo, de quem, de onde, quando e como as mensagens são originadas. Uma diversidade de informações estaria disponível se técnicas de *data mining* pudessem descobrir por que esses *e-mails* foram enviados. Um dos pontos comuns de interação entre o cliente e a empresa é a central de atendimento. A HP está impressionada com os *insights* de negócios que poderiam ser escolhidos de comunicações, como documentos do Word, *e-mails* e outras fontes. A combinação de dados estruturados e não estruturados pode criar um enorme potencial para empresas descobrirem valiosos *insights* de negócios.

SISTEMA

As ferramentas convencionais que a HP usava anteriormente não conseguiam relatar informações úteis provenientes de comunicações relacionadas ao cliente. Por isso, a HP usou o Text Miner do SAS Institute para descobrir *insights* analíticos de dados relacionados ao cliente em aplicações da central de atendimento e, então, padronizou esses *insights*. A HP implementou o Text Miner para combinar dados estruturados e dados de texto a fim de produzir um conjunto híbrido de dados estruturados/não estruturados que estão armazenados no banco de dados Microsoft SQL Server com um mecanismo OLAP de serviços de análise. Hoje, o sistema contém 300.000 documentos de texto e tem aproximadamente 50 *gigabytes* de extensão, abrangendo um período de 18 meses e 3 centrais de atendimento.

A HP implementou o Executive Viewer, uma ferramenta baseada na Web desenvolvida pela Temtec (**temtec.com**) que permite à empresa ampliar os cubos OLAP com modelagem preditiva, pontuações de fidelidade e diferenciais para o cliente criados pelo SAS Enterprise Miner.

PROCESSO

Muitos conceitos, como produtos usados, frequência de ligação e questões gerais do cliente, são usados para agregar dados de textos; o resultado é consolidado em grupos de texto probabilísticos. Então, esse grupo consolidado é combinado com os dados estruturados de fornecedores externos. Agora, a HP pode combinar e analisar dados estruturados como, por exemplo, rendimentos com pedidos, atitudes e necessidades do cliente.

A análise de texto é desafiadora devido a dimensionalidade e dispersão de dados. Bancos de dados

de cliente distintos contêm diferentes informações estruturadas que poderiam ser integradas sem muita dificuldade. O desafio está na combinação de dados estruturados com dados não estruturados provenientes do texto. O SAS Text Miner usa uma técnica chamada *decomposição em valores singulares*. O *software* de *text mining* inclui um dicionário pré-desenvolvido de listas de palavras e sinônimos; é uma tarefa desgastante para a empresa personalizar as informações de texto geradas nos seus ambientes de negócios. Dados de texto estão disponíveis em várias fontes fora do campo do *data warehousing* tradicional. Alguns dos maiores desafios que o SAS Text Miner enfrenta são as atividades do cliente no site da HP e também encontrar *insights* nos negócios dos clientes da HP.

Além de ser usado como a principal aplicação de *text mining*, o SAS Text Miner poderia ser usado de modo proativo no site do cliente para gerar *insights* nas necessidades do cliente que poderiam ser satisfeitas pela HP. A ferramenta também poderia ser usada para analisar múltiplos fornecedores com diversos números e descrições em texto.

RESULTADOS

SAS Text Miner foi capaz de desenvolver, com sucesso, definições padrão de dados e modelos para classificação de produto com mais de 80% de precisão. Atualmente, o sistema está sendo usado para dar suporte à HP na contribuição ao faturamento através de venda cruzada melhorada, *marketing* direcionado, retenção do cliente e melhor antecipação das necessidades do cliente. A informação gerada de dados estruturados/não estruturados agora dá suporte a múltiplos usuários de negócios em vários departamentos.

Fontes: M. Hammond, "BI Case Study: What's in a Word? For Hewlett-Packard, It's Customer Insight." *Business Intelligence Journal*, Vol. 9, No. 3, Summer 2004, p. 48–51; and B. Beal, "Text Mining: A Golden Opportunity for HP", *SearchCRM.com*, June 6, 2005, searchdatamanagement.techtarget.com/originalContent/0,289142,sid91_gci1136611,00.html (acessado em abril de 2006).

QUESTÕES SOBRE O CASO

1. Qual é a aplicação prática de *text mining*?
2. Como você acha que as técnicas de *text mining* poderiam ser usadas em outras empresas?
3. Quais foram os desafios em *text mining* da HP? Como eles foram superados?
4. Na sua opinião, em quais outras áreas a HP pode usar o *text mining*?

Referências

- Bhandari, I., E. Colet, J. Parker, Z. Pines, R. Pratap, and K. Ramanujam. (1997). "Advanced Scout: *Data Mining* and Knowledge Discovery in NBA Data". *Data Mining and Knowledge Discovery*, Vol. 1, No. 1, pp. 121–125.
- Buck, N. (December 2000/January 2001). "Eureka! Knowledge Discovery", *Software Magazine*.
- Cao, J., J.M. Crews, M. Lin, J.K. Burgoon, and J.F. Nuna-maker, Jr. (2003, June 2–3). "Designing Agent99 Trainer: A Learner-Centered, Web-Based Training System for Deception Detection". *Proceedings of the First NSF/NIJ Symposium on Intelligence and Security Informatics (ISI 2003)*, Tucson, AZ, pp. 358–365.
- Cavoukian, A. (1998). *Data Mining: Staking A Claim on Your Privacy*, Report by Information and Privacy Commissioner, Ontario, Canada.
- ipc.on.ca/index.asp?navid=46&fid1=342 (acessado em agosto de 2006).
- Chan, P.K., W. Phan, A. Prodromidis, and S. Stolfo. (1999, November/December). "Distributed *Data Mining* in Credit Card Fraud Detection", *IEEE Intelligent Systems*.
- Copeland, L. (2001, October 22). "Developers Approach Extreme Programming with Caution". *Computerworld*.
- Davenport, T.H. (2006, January). "Competing on Analytics". *Harvard Business Review*.
- Dunham, M. (2003). *Data Mining: Introductory and Advanced Topics*. Upper Saddle River, NJ: Prentice Hall.
- Ellingsworth, M., and D. Sullivan. (2003, July). "Text Mining Improves Business Intelligence and Predictive Modeling in Insurance". *DM Review*.
- Etzioni, O. (1996). "The World Wide Web: Quagmire or Gold Mine". *Communications of the ACM*, Vol. 39, No. 11.
- Fadlalla, A., and C-H. Lin. (2001, August). "An Analysis of the Applications of Neural Networks in Finance". *Interfaces*.
- Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth. (1996). "From Knowledge Discovery in Databases". *AI Magazine*, Vol. 17, No. 3.
- Finlay, D. (2001, April 15). "Real-Time Intelligence Scores Over OLAP", *Software Development Times*.
- Grimes, S. (2001, June 13). "Mining a Demographic Mother Lode", *Intelligent Enterprise*.
- Glymour, C., D. Madigan, D. Pregibon, and P. Smyth. (1997). "Statistical Themes and Lessons for *Data Mining*". *Data Mining and Knowledge Discovery*, Vol. 1, No. 1.
- Groth, R. (1998). *Data Mining: A Hands-on Approach for Business Professionals*. Upper Saddle River, NJ: Prentice Hall.
- Hoffman, T. (1999, April 19). "Insurers Mine for Age-Appropriate Offering". *Computer World*.
- Hoffman, T. (1998, December 7). "Banks Turn to IT to Reclaim Most Profitable Customers". *Computer World*.
- Kleinberg, K.M. (1998, January 25–27). "Authoritative Sources in Hyperlinked Environment", *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, San Francisco.
- Morris, S.A., and K.W. Boyack. (2005, July 24–28). "Visualizing 60 Years of Anthrax Research". *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics*, Stockholm, Sweden.
- Nemati, H.R., and C.D. Barko. (2001, Winter). "Issues in Organizational *Data Mining*: A Survey of Current Practices". *Journal of Data Warehousing*, Vol. 6, No. 1.
- Pallatto, J. (2002, February). "Business Tools Get Smart". *Internet World*.
- Palshikar, G.K. (2001, April 16). "Matching Patterns". *Intelligent Enterprise*.
- Quinlan, J.R. (1986, March). "Induction of Decision Trees". *Machine Learning*.
- Statsoft.com. (2006). "*Data Mining* Techniques". statsoft.com/textbook/stdatmin.html (acessado em agosto de 2006).
- Srivastava J., P. Desikan and V. Kumar (2002, November 1–3), "Web Mining: Accomplishments and Future Directions", *Proceedings of the US National Science Foundation Workshop on Next-Generation Data Mining (NGDM)*, National Science Foundation, 2002, Baltimore, MD. csee.umbc.edu/~kolari1/Mining/papers/srivastava.pdf (acessado em agosto de 2006).
- Zhou, Y., E. Reid, J. Qin, H. Chen, and G. Lai. (2005, September/October). "U.S. Domestic Extremist Groups on the Web: Link and Content Analysis". *IEEE Intelligent Systems*.



Business Performance Management (BPM)

Objetivos de aprendizado

- ◆ Entender a natureza abrangente do *business performance management* (BPM)
- ◆ Entender o processo de retorno ligando a estratégia à execução
- ◆ Descrever algumas das melhores práticas de planejamento e relatório de gerenciamento
- ◆ Descrever a diferença entre gestão de desempenho e medidas
- ◆ Entender o papel das metodologias no BPM
- ◆ Descrever os elementos básicos das metodologias *balanced scorecard* e Six Sigma
- ◆ Descrever as diferenças entre *scorecard* e painéis
- ◆ Entender alguns dos pontos básicos de projeto de painéis
- ◆ Entender os usos potenciais de monitoramento de atividades de negócios (BAM)

O *business performance management* (BPM) representa a próxima fase na evolução de sistemas de suporte à decisão (DSS), sistemas de informações empresariais e *business intelligence* (BI). Faz pelo menos 25 anos que este sistema está em processo de evolução. Como suporte a decisões, o BPM é mais que uma ferramenta de tecnologia. É um conjunto integrado de processos, metodologias, métricas e aplicações projetadas para impelir o desempenho geral financeiro e operacional de uma empresa. Ajuda as empresas a converterem suas estratégias e objetivos em planos, monitorar o desempenho em relação a esses planos, analisar variações entre resultados reais e resultados planejados, e ajustar seus objetivos e ações em resposta a essa análise.

O capítulo examina os processos, metodologias, métricas e sistemas subjacentes ao BPM. Devido ao fato de o BPM se distinguir do DSS e BI por sua concentração em estratégia e objetivos, o capítulo se inicia com uma exploração das noções de estratégia empresarial e execução e a lacuna que geralmente existe entre eles. As seções específicas são:

- 5.1 Vinheta de abertura: Cisco e o fechamento virtual
- 5.2 Visão geral do gerenciamento do desempenho de negócios ou *business performance management* (BPM)
- 5.3 Estratégia: aonde queremos ir?
- 5.4 Plano: como chegamos lá?
- 5.5 Monitoração: como estamos fazendo?

- 5.6 Ação e ajuste: o que precisamos fazer de forma diferente?
- 5.7 Medida de desempenho
- 5.8 Metodologias de BPM
- 5.9 Arquitetura e aplicações de BPM
- 5.10 Painéis de desempenho
- 5.11 Monitoramento de Atividades de Negócios ou *business activities monitoring* (BAM)

5.1

VINHETA DE ABERTURA: CISCO E O FECHAMENTO VIRTUAL

A Cisco Systems se refere como a “líder mundial em redes para a Internet”. Ela oferece soluções de rede – *hardware*, *software* e serviços – que conectam os dispositivos de computação e redes que compõem a Internet e a maioria das redes corporativas, educacionais e governamentais ao redor do mundo. A Cisco foi fundada em 1984. Na maior parte da sua história, ela tem sido um sucesso financeiro. Por exemplo, em 2005, a Cisco relatou rendas de mais de US\$ 25 bilhões. A capitalização de mercado da empresa (i.e., o valor de suas impressionantes ações) é de acima de US\$ 100 bilhões, o que torna a Cisco uma das maiores empresas de computação do mundo.

Problema

Além de ser uma pioneira e líder de mercado no setor de computação, a Cisco também é uma pioneira em relatórios de gerenciamento e financeiros. Durante o *boom* da Internet e do “.com”, as rendas da Cisco cresceram em uma taxa composta anual de mais de 60%. Grande parte do crescimento foi estimulado pela aquisição de empresas menores. O diretor financeiro chefe, Larry Carter, estava preocupado que o sistema de relatório financeiro da Cisco não acompanhasse o rápido crescimento. Para resolver o problema, ele decidiu instituir um processo chamado “fechamento virtual”, que forneceria a capacidade de determinar o estado financeiro da empresa com o aviso de uma hora e de disseminar as informações instantaneamente através da *intranet* da empresa. Para realizar isso, a Cisco teria de divulgar todas as suas transações financeiras (p.ex., inventário, ordenados, contas) quase em tempo real. Nesse período, levava duas semanas para a Cisco fechar seus registros financeiros. Hoje em dia, ainda leva duas ou mais semanas para muitas empresas realizarem essa tarefa. Em 1999, o departamento financeiro da Cisco obteve sucesso em instituir o novo processo.

O novo processo da Cisco foi anunciado pela imprensa e pelos analistas de mercado. O sistema funcionou bem por alguns anos, até a explosão das “.com”, em 2001. Nesse ano, quando as companhias de telecomunicação começaram a falir, a demanda pelos produtos da Cisco caiu dramaticamente. A Cisco ficou com estoque de inventário em excesso de uma hora para outra. No terceiro trimestre daquele ano, a empresa relatou perdas de quase US\$ 3 bilhões. Suas ações foram de aproximadamente de US\$ 80 por ação a cerca de US\$ 8 por ação. O que deu errado no fechamento virtual? O sistema deveria fornecer à Cisco a capacidade de reagir rapidamente às condições em mudança do mercado. Uma explicação fornecida pelo CEO John Chambers foi a de que o fechamento virtual permitiu que a empresa rastreasse informações financeiras e outras informações operacionais de forma diária, mas não permitiu que a Cisco predissesse o futuro, especialmente as tendências macroeconômicas. Qualquer que seja a explicação, a Cisco não conseguiu evitar os declínios econômicos dramáticos que se iniciaram no início de 2001, mesmo que tivesse alocado um sistema de medidas de altíssimo nível.

Solução

Diferentemente de muitas outras empresas, a Cisco pode suportar o que Chambers denominou “tempestade de um século”. Foi necessário uma combinação de analistas, pensamento estratégico e execução tática. Não havia problemas no sistema de medidas. De fato, alguns analistas sugeriram que o dano à Cisco seria maior se o fechamento virtual não estivesse ativo. Com base em sua análise, a gerência sênior da Cisco pensou que o mercado de telecomunicações estava encolhendo, mas que não desapareceria. Era simplesmente uma queda cíclica, e não uma mudança permanente. Não havia necessidade de trocar produtos ou mudar dramaticamente o modelo de negócios da Cisco. Em vez disso, a empresa precisava melhorar muito a produtividade. Ela o fez através de uma série de etapas. A Cisco reduziu drasticamente em US\$ 2,5 bilhões seu estoque. Depois, cortou despesas pela redução de pessoal em 20.000 funcionários. Por fim, a empresa reduziu o número de modelos para cada grupo de produto e começou a usar seu desenvolvimento interno no lugar de depender da aquisição de produtos para aumentar suas ofertas.

Naturalmente, é fácil melhorar a lucratividade cortando custos. Mas isso pode funcionar por apenas um curto período de tempo. No final das contas, uma empresa tem que descobrir maneiras de fazer crescer seu negócio se ela quiser sobreviver com o correr do tempo. Felizmente, a Cisco também ajustou sua estratégia de crescimento ao mesmo tempo em que lidava com custos e produtividade. Ela acreditava que podia aumentar seu negócio e especialidade técnica para entrar em novos mercados. Em 2003, a Cisco entrou em seis novas áreas de produtos, incluindo redes ópticas, redes locais sem fio (LANs), segurança de rede, telefonia IP, redes de armazenamento, e rede de consumidor/doméstica. Chambers expressou sua crença de que cada uma dessas novas áreas tinha o potencial de criar US\$ 1 bilhão em receitas.

Resultados

Em 2005, a receita da Cisco era de US\$ 25 bilhões, com uma receita líquida de perto de US\$ 6 bilhões. Isso representou um aumento de 12% nas vendas em relação ao ano anterior. O crescimento da empresa continua a ser abastecido por aquisições, como a compra, em 2003, da empresa de rede doméstica e de pequenas empresas Linksys, bem como seus investimentos internos em novas áreas de produção. Conforme predisse a Cisco, as novas áreas de produtos se tornaram oportunidades de bilhões de dólares. Por exemplo, as ofertas de segurança de rede da Cisco e de telefonia IP geram, agora, cerca de US\$ 1 bilhão cada ao ano.

A Cisco está liderando a transição para os ambientes de rede inteligentes. Seu sucesso resultou do planejamento contínuo de três a cinco anos à frente de importantes transições de mercado. Conforme disse Chambers, “Historicamente, as transições de que fizemos parte têm sido tecnológicas; elas foram relativamente ordenadas e previsíveis. Agora, todavia, estamos vendo uma transição de maior alcance e mais dramática. Logo, nossa capacidade de predizer aonde irá o mercado é mais crítica ainda e oferece um potencial muito maior de importação para a empresa.”

Fontes: compilado de A. Hartman, *Ruthless Execution*, Prentice Hall, Upper Saddle River, NJ, 2004; G. Moore, *Dealing with Darwin*, Penguin, New York, 2005; and *Corporate Overview*, newsroom.cisco.com/dlls/company_overview.html (acessado em abril de 2006).

Questões sobre a vinheta de abertura

1. O que é o sistema de fechamento virtual?
2. Que explicação John Chambers, CEO da Cisco, dá para a queda da empresa em 2001?
3. Qual foi a resposta estratégica da Cisco para sua queda financeira em 2001?

O que podemos aprender com essa vinheta?

O fechamento virtual da Cisco permitiu que a empresa monitorasse indicadores essenciais de importância operacional ou tática. O que não fez, foi fornecer à empresa os meios de moldar certas acepções estratégicas críticas, como o crescimento de mercado ou a falta de crescimento. Como a Cisco aprendeu, o desempenho eficaz dos negócios exige que uma empresa modelar e monitorar não apenas suas táticas, mas também suas estratégias e as pressuposições sobre as quais essas estratégias são construídas.

5.2

VISÃO GERAL DO *BUSINESS PERFORMANCE MANAGEMENT* (BPM)

Conforme exemplifica a experiência da Cisco, é difícil para um negócio ou uma empresa alinhar suas estratégias, planos, sistemas analíticos e ações de maneira que assegurem o desempenho bem-sucedido. É a esse alinhamento que a gestão de desempenho, em oposição à medida de desempenho, se refere.

BPM definido

Na literatura de negócios e comércio, a gestão de desempenho tem vários nomes, o que inclui BPM, gestão de desempenho corporativo (CPM), gestão de desempenho empresarial (EPM) e gestão estratégica da empresa (SEM). Embora diferentes termos sejam utilizados, todos eles significam essencialmente a mesma coisa. Por exemplo, SEM é um termo que a SAP (**sap.com**) utiliza, e foi definido como “as técnicas de gerenciamento, métricas e ferramentas relacionadas (como o *software* de computador) projetados para darem assistência às empresas a tomar decisões estratégicas de alto nível” (Wikipedia.org, 2006). De forma semelhante, a firma de análise de mercado Gartner (**gartner.com**) definiu CPM como “um termo-guarda-chuva que cobre os processos, metodologias, métricas e tecnologias para que as empresas meçam, monitorem, e gerenciem o desempenho de negócios” (Buytendijk, 2004). Finalmente, o BPM Standards Group (2005) (**bpmstandardsgroup.com**) definiu o *business performance management* (BPM) como “uma estrutura para organizar, automatizar e analisar as metodologias de negócios, métricas, processos e sistemas, de modo a impelir o desempenho geral da empresa. O ajuda as empresas a converterem um conjunto unificado de objetivos em planos, monitorar a execução, e fornecer um *insight* crítico para melhorar o desempenho financeiro e operacional.” Neste capítulo, o termo *BPM* é utilizado, em vez dos outros termos, pois é o termo utilizado pelo grupo de padrões.

BPM e BI comparados

O BPM é um resultado da BI e incorpora muitas de suas tecnologias, aplicações e técnicas. Alguns críticos do BPM se perguntam por que a BI não pode fornecer o *insight* necessário para melhorar o desempenho geral dos negócios. De um ponto de vista teórico, pode. De um ponto de vista prático, não forneceu. Por exemplo, muitas empresas concentraram suas iniciativas de BI em questões departamentais, mais do que em problemas empresariais. A Tabela 5.1 dá detalhes sobre alguns dos meios nos quais as empresas usaram tradicionalmente BI *versus* a maneira pela qual a BI deve ser implementada para dar apoio ao BPM.

O BPM não é novo. Cada empresa de certo tamanho tem processos ativados (p. ex., orçamentos, planos detalhados, execução, medida) que retornam ao plano geral. O que o conceito de BPM adiciona ao que as empresas têm feito por anos é a integração desses processos, metodologias, métricas e sistemas. O BPM é uma estratégia empresarial que busca prevenir que as empresas otimizem negócios locais às expensas do desempenho geral corporativo. Não é um projeto “um por vez”, mas um processo contínuo – parte do trabalho diário de gerentes.

Tabela 5.1 Diferença entre BI tradicional e BI para BPM

Fator	BI tradicional	BI para BPM
Escala	Departamental	Empresarial
Concentração	Histórico	Oportuno
Decisões	Estratégico e tático	Estratégico, tático e operacional
Usuários	Analistas	Todos
Orientação	Reativo	Proativo
Processo	Dimensão ilimitada	Ciclo fechado
Medidas	Métricas	Indicadores principais de desempenho
Vistas	Genérico	Personalizado
Visuais	Tabelas/diagramas	Painéis/placares
Colaboração	Informal	Embutido
Interação	Puxar (perguntas <i>ad hoc</i>)	Empurrar (alertas)
Análise	Tendências	Exceções
Dados	Somente numérico	Numérico, texto, etc.

Fonte: W. Eckerson, *Best Practices in Business Performance Management: Business and Technical Strategies*, March 2004, tdwi.org/display.aspx?id=6972 (acessado em outubro de 2006).

Resumo dos processos de BPM

O BPM engloba um conjunto de ciclo fechado de processos que liga estratégia a execução, de forma a otimizar o desempenho dos negócios (ver Figura 5.1). O ciclo implica que o desempenho perfeito é alcançado pela definição de metas e objetivos (i.e., fazer estratégias), estabelecer iniciativas e planos para alcançar essas metas (i.e., planejar), monitorar o desempenho real em relação

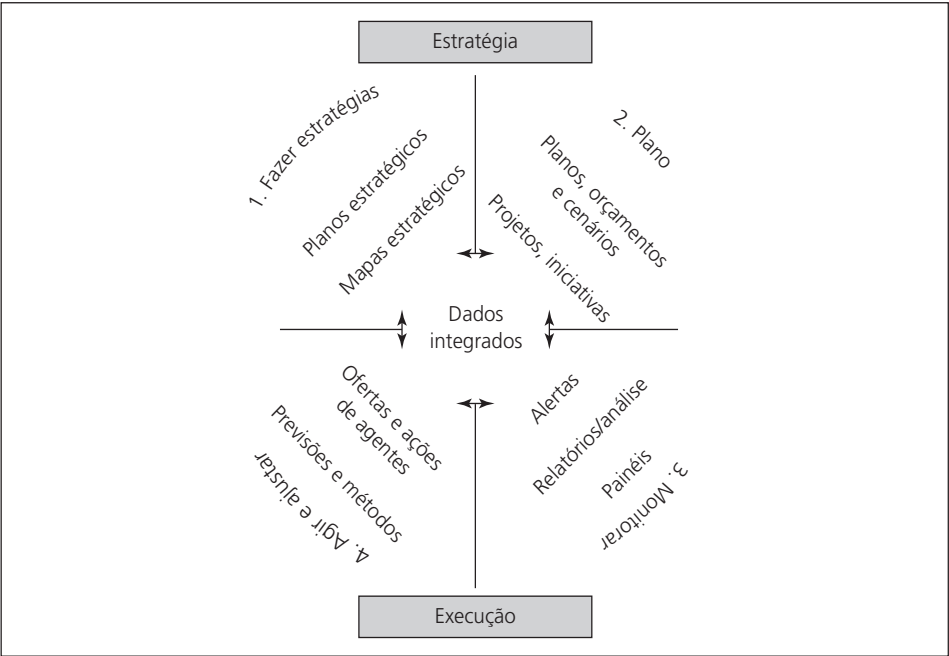


Figura 5.1 Processos do ciclo fechado de BPM.

Fonte: W. Eckerson. *Performance Dashboard*, Wiley. Hoboken, NJ, 2006.

às metas e objetivos (i.e., monitorar) e tomar ações corretivas (i.e., agir e ajustar). Nas Seções 5.3 até 5.6, cada um desses principais processos é examinado em detalhes.

Questões de revisão da Seção 5.2

1. Defina *BPM*.
2. Resuma as maneiras pelas quais a BI para o BPM difere da BI tradicional.
3. Liste os principais processos de BPM.

5.3

ESTRATÉGIA: AONDE QUEREMOS IR?

Durante a era do “.com” muito gerentes e conhecedores de negócios se convenceram que a estratégia era supérflua. Em suas mentes, a complexidade e andamento da mudança no ambiente de negócios tornaria qualquer estratégia obsoleta em um curto período de tempo. Nas palavras de um desses conhecedores (Tarlow e Tarlow, 2002),

A maior parte da literatura de negócios padrão ainda depende da idéia de que temos que definir nossas metas, definir prioridades, desenvolver nossas estratégias, gerenciar nossas produções e avaliar nossos impactos. Posso assegurar-lhes de que se operarem dessa maneira, alguém já venceu você na reta final. Não se pode planejar rápido o suficiente. Precisamos de comportamentos que são muito mais ousados e afinados à natureza exclusiva do nosso tempo.

Conforme Michael Porter (2001) definiu em seu artigo vencedor do McKinsey Award, no *Harvard Business Review*, “Strategy and the Internet”, muitas empresas seguem esse tipo de conselho. Durante esse período, “as empresas “.com” e as estabelecidas violaram quase todos os preceitos da boa estratégia” – caçando clientes indiscriminadamente, perseguindo rendas indiretas, oferecendo qualquer produto ou serviço concebível, em vez de se concentrar em lucros, focando-se em fornecer valor e fazendo *trade-offs* competitivos. Os resultados de seguir esse conselho são bem conhecidos agora. Em março de 2000, havia 378 de capital aberto na Internet. Sua capitalização de mercado era de US\$ 1,5 trilhão. Entretanto, elas combinaram as vendas anuais de apenas US\$ 40 bilhões e 87% nunca demonstraram um lucro trimestral. Até o verão de 2001, sua capitalização de mercado caiu 75%, elas desempregaram 31.000 funcionários e 130 fecharam as suas portas. A nova economia estava terminada quase antes de começar.

Hoje em dia, o pêndulo oscilou para trás. As empresas reconhecem o valor do planejamento estratégico, bem como a necessidade de excelência operacional ou de execução. Sem metas ou objetivos específicos, é difícil avaliar cursos alternativos de ação. Sem prioridades específicas, não há maneira de determinar como alocar recursos entre as alternativas selecionadas. Sem planos, não há maneiras de guiar as ações entre aqueles que estão trabalhando nas alternativas. Sem análise e avaliação, não há maneira de determinar quais das oportunidades estão sendo bem-sucedidas ou falhando. Metas, objetivos, prioridades, planos e pensamento crítico, todos formam parte de uma estratégia bem-definida.

Planejamento estratégico

O termo *estratégia* tem muitas definições. Para aumentar a confusão, também é usado com frequência em combinação com uma variedade de outros termos, como visão estratégica e concentração estratégica. Independentemente das diferenças de significado, todos enfrentam a questão “Aonde queremos ir no futuro?” Para a maioria das empresas, a resposta a essa questão é fornecida em um plano estratégico. Você pode pensar em um plano estratégico como um

mapa, detalhando um curso de ações para mover uma empresa de seu estado atual para a sua visão futura.

Tipicamente, os planos estratégicos se iniciam no topo e no início com uma visão da amplitude da empresa. A partir daí, planos estratégicos são criados para as unidades de negócios da empresa, ou unidades funcionais. As seguintes tarefas são um tanto comuns para o processo de planejamento estratégico, independentemente do nível no qual o planejamento é feito – pela empresa toda, em unidades de negócios, ou em unidades funcionais (Wade e Recardo, 2001):

1. *Conduzir uma análise de situação atual.* Essa análise examina a situação atual da empresa (“Onde estamos?”) e estabelece uma linha de base, bem como principais tendência, para o desempenho financeiro e o desempenho operacional.
2. *Determinação do horizonte de planejamento.* Tradicionalmente, as empresas produzem planos com base anual, com o horizonte de planejamento operando em três ou cinco anos. Em grande parte, o horizonte de tempo é determinado pela volatilidade e previsibilidade do mercado, ciclos de vida de produtos, o tamanho da empresa, a taxa de inovação tecnológica e a intensidade capital da indústria. Quanto mais volátil, menos previsível, menores os ciclos de vida, menor a empresa, maior a taxa de inovação, e menor a intensidade capital, menor o horizonte de planejamento.
3. *Conduzir uma varredura de ambiente.* Uma varredura de ambiente é um julgamento padrão de forças, fraquezas, oportunidades e ameaças (SWOT) da empresa. Identifica e prioriza os principais fatores de cliente, mercado, concorrentes, governo, demográficos, acionistas, e de indústria potencialmente ou realmente que afetam a empresa.
4. *Identificar fatores críticos de sucesso.* Os **fatores críticos de sucesso (FCS)** delineiam aquelas coisas nas quais a empresa deve se sobressair para ter sucesso no seu espaço de mercado. Para uma empresa concentrada em produto, qualidade e inovação de produto são exemplos de FCS. Para um fornecedor de baixo custo como o Wal-Mart, as capacidades de distribuição são um FCS.
5. *Análise de compleição de uma lacuna (análise de gap).* Como a varredura de ambiente, uma análise de lacuna é utilizada para identificar e priorizar as forças e fraquezas internas no processo de uma empresa, suas estruturas, tecnologias e aplicações. As lacunas refletem o que a estratégia realmente exige e o que a empresa realmente alcança.
6. *Criar uma visão estratégica.* A **visão estratégica** de uma empresa fornece uma figura ou imagem mental de como uma empresa deveria parecer no futuro – a mudança em seus produtos e mercados. Geralmente, a visão está expressa em termos do seu foco estratégico e identifica o estado atual e o estado desejado de uma empresa.
7. *Desenvolver uma estratégia de negócios.* O desafio nesta etapa é produzir uma estratégia que é baseada nos dados e informações das etapas anteriores e seja consistente com a visão estratégica da empresa. O senso comum nos diz que a estratégia precisa explorar as forças da empresa, tomar vantagem das suas oportunidades, lidar com fraquezas e responder a ameaças. A empresa precisa assegurar que sua estratégia está internamente consistente, que a sua cultura está alinhada com a estratégia e que há suficientes recursos e capital para implementá-la estratégia.
8. *Identificar objetivos e metas estratégicas.* Um plano estratégico que não consegue fornecer direções claras para o processo de planejamento operacional e financeiro é incompleto. Antes que um plano operacional ou financeiro possa ser estabelecido, objetivos estratégicos devem ser estabelecidos e refinados em metas ou alvos bem-definidos. Um **objetivo estratégico** é uma declaração ampla ou o curso geral de uma ação que prescreve direções com alvo para uma empresa. Antes que um objetivo estratégico possa ser ligado a um plano operacional ou plano financeiro, deveria ser convertido em uma meta bem-definida ou alvo. Uma **meta estratégica**

é uma quantificação de um objetivo para um período designado de tempo. Por exemplo, se uma empresa tem um objetivo de melhorar os rendimentos de ativos (ROA) ou aumentar a lucratividade geral, esses objetivos precisam ser transformados em alvos quantificados (p.ex., um aumento no ROA de 10 a 15% ou um aumento na margem de lucro de 5 a 7%) antes que a empresa possa começar a detalhar os planos operacionais necessários para alcançar esses alvos. No mesmo espírito, quando a Cisco declarou que ia expandir sua oferta de produtos em seis áreas de tecnologia avançada, esse objetivo precisou ser transformado em um alvo bem-definido, como fazer crescer cada área de produto de tecnologia avançada em habituais de US\$ 1 bilhão por ano, antes que a empresa pudesse tornar os investimentos de produtos e decisões de aquisições. As metas e alvos estratégicos guiam a execução operacional e permitem que o progresso seja rastreado em relação aos objetivos gerais.

A lacuna estratégica

Uma coisa é criar uma estratégia de longo prazo; outra, executá-la. Um estudo da revista *Fortune* de 1999 descobriu que 70% das falhas de CEOs eram o resultado de execução ruim, ao invés de estratégias ruins (Charan e Colvin, 1999). Niven (2005) apontou quatro fontes para a lacuna entre estratégia e execução:

1. *Visão.* Uma citação do filme *Rebeldia Indomável*, “O que temos aqui é uma falha de comunicação”, se aplica à visão estratégica. Em muitas empresas, somente uma percentagem muito pequena (cerca de 5%) dos funcionários entende a estratégia da organização. Sem esse conhecimento, é difícil, se não impossível, para os funcionários tomarem decisões e agirem de acordo com o plano estratégico.
2. *Pessoas.* Ligar pagamento a desempenho é importante para a execução bem-sucedida. Entretanto, planos de incentivo são frequentemente ligados a resultados financeiros de curto prazo, não ao plano estratégico ou mesmo às iniciativas estratégicas articuladas no plano operacional. Maximizar ganhos de curto prazo leva a tomadas de decisão menos que racionais.
3. *Gerenciamento.* O gerenciamento geralmente gasta tempo na periferia dos problemas, em vez de se concentrar nos elementos principais. Podem-se gastar horas debatendo itens de linha em um orçamento, com pouca atenção dada à estratégia, a ligação do plano financeiro à estratégia ou as pressuposições subjacentes à ligação.
4. *Recursos.* Em tempos recentes, críticos (p. ex., o *Beyond Budgeting Round Table*, [bbrrt.org]), questionaram a necessidade de orçamentos e o processo de orçamento. A menos que iniciativas estratégicas sejam fundamentadas e tenham recursos apropriados, sua falha é quase que certa.

Questões de revisão da Seção 5.3

1. Por que uma empresa precisa de uma estratégia bem-formulada?
2. Quais são as tarefas básicas no processo de planejamento estratégico?
3. Quais são algumas das fontes da lacuna entre formular uma estratégia e realmente executar a estratégia?

5.4

PLANO: COMO CHEGAREMOS LÁ?

Quando os gerentes operacionais sabem e entendem o *que* (i.e., os objetivos e metas organizacionais), eles podem vir com o *como* (i.e., planos detalhados operacionais e financeiros). Planos financeiros e operacionais respondem a duas questões: que táticas e iniciativas serão perseguidas para atender aos alvos de desempenho estabelecidos pelo plano estratégico? Quais são os resultados financeiros esperados da execução da tática?

Planejamento operacional

Um **plano operacional** converte uma estratégia operacional e metas de uma empresa em um conjunto de táticas e iniciativas bem-definidas, exigências de recursos e resultados esperados para um período de tempo futuro, geralmente, mas nem sempre, um ano. Essencialmente, um plano operacional é como um plano de projeto que é designado para assegurar que a estratégia de uma empresa seja realizada. A maioria dos planos operacionais englobam um portfólio de táticas e iniciativas. A chave para o planejamento operacional bem-sucedido é a integração. A estratégia impele táticas, e táticas impelem resultados. Basicamente, as táticas e iniciativas definidas em um plano operacional devem ser ligadas diretamente a objetivos e alvos principais no plano estratégico. Se não houver elo entre uma tática individual e um ou mais objetivos estratégicos ou alvos, o gerenciamento deve questionar se a tática e suas iniciativas associadas são realmente necessárias. As metodologias de BPM debatidas na Seção 5.8 são feitas para assegurar que esses elos existem.

O planejamento operacional pode ser centrado em táticas ou orçamentos (ver Axson, 2003). Em um plano *centrado em táticas*, as táticas são estabelecidas para atender aos objetivos e metas estabelecidas no plano estratégico. De modo inverso, em um plano *centrado em orçamentos*, um plano financeiro ou orçamento é estabelecido de forma que some aos valores financeiros finais. As empresas que usam as melhores práticas usam o planejamento operacional centrado em táticas. Isso significa que elas começam o processo de planejamento operacional pela definição de táticas e iniciativas alternativas que podem ser usadas para alcançar um alvo em especial. Por exemplo, se uma empresa está com o alvo em um crescimento de 10% na margem de lucro (i.e., a razão da diferença entre renda e despesas dividida pela receita), a empresa primeiro determinará se planeja aumentar a margem pelo aumento de receita, pela redução de despesas ou alguma combinação de ambos. Se ela se concentrar nas receitas, então a questão se tornará se planeja entrar em novos mercados ou aumentar as vendas em mercados existentes, se planeja melhorar produtos existentes ou introduzir novos produtos ou alguma combinação desses. Os cenários alternativos e iniciativas associadas devem ser pesados em termos de seu risco geral, exigências de recursos, e viabilidade financeira. O Arquivo Online W5.1 debate as etapas utilizadas em tomar decisões entre os vários cenários.

Planejamento e orçamento financeiro

Na maioria das empresas, há uma tendência para a redução de recursos. Se isso não acontecesse, as empresas poderiam simplesmente jogar pessoas e dinheiro nas suas oportunidades e problemas e sobrepujar a concorrência. Dada a raridade dos recursos, uma empresa precisa pôr seu dinheiro e pessoal aonde suas estratégias e táticas vinculadas estejam. Os objetivos estratégicos de uma empresa e métricas essenciais deveriam servir de impelidores do topo para a base para a alocação dos bens tangíveis e não-tangíveis de uma empresa. Enquanto as operações contínuas necessitam claramente de apoio, os recursos essenciais deveriam ser atribuídos aos mais importantes programas e as prioridades estratégicas. A maioria das empresas usam seus orçamentos e programas de compensação para alocar recursos. Por implicação, ambos precisam ser cuidadosamente alinhados com os objetivos estratégicos e táticos da empresa, de modo a alcançar sucesso estratégico.

A melhor maneira de uma empresa alcançar esse alinhamento é basear seu plano financeiro em seu plano operacional, ou, mais diretamente, alocar e orçar seus recursos contra táticas e iniciativas específicas. Por exemplo, se uma das táticas for desenvolver um novo canal de vendas, receitas orçamentadas e custos devem ser atribuídos ao canal, no lugar de simplesmente ter custos atribuídos a unidades funcionais em particular, como *marketing* e P&D. Sem esse tipo de planejamento de recursos tático, não há maneira de medir o sucesso dessas táticas e, portanto, da estraté-

gia. O tipo de ligação ajuda as empresas a evitar o problema de cortes orçamentários “aleatórios” que afeta inadvertidamente as estratégias associadas. O orçamento baseado em táticas assegura que o elo entre os itens de linha de orçamento particulares e as táticas particulares ou iniciativas seja bem-estabelecido e bem-conhecido.

O processo de planejamento financeiro e orçamentário tem uma estrutura lógica que se inicia tipicamente com aquelas táticas que geram alguma forma de receita ou rendimento. Em empresas que vendem bens ou serviços, a capacidade de gerar receita se baseia ou na capacidade de produzir diretamente produtos ou serviços ou adquirir a quantidade certa de bens e serviços para vender. Após um valor de receita ter sido estabelecido, os custos associados para fornecer o nível de receita podem ser gerados. Com muita frequência, isso demanda dados e táticas de muitos departamentos. Isso significa que o processo tem que ser colaborativo e que as dependências entre as funções devem ser claramente comunicadas e compreendidas. Além da entrada colaborativa, a empresa também precisa adicionar vários custos gerais, bem como os custos do capital necessário. Essas informações, uma vez consolidadas, mostram os custos por tática, bem como as exigências de dinheiro e fundos exigidas para pôr o plano em operação.

Questões de revisão da Seção 5.2

1. Qual é a meta do planejamento operacional?
2. O que é o planejamento centrado em táticas? O que é o planejamento centrado em orçamento?
3. Qual é a meta primária de um plano financeiro?

5.5

MONITORAÇÃO: COMO ESTAMOS FAZENDO?

Quando o plano operacional e os planos financeiros estão a caminho, é imperativo que o desempenho da empresa seja monitorado. Uma estrutura abrangente para monitorar o desempenho deveria lidar com dois problemas principais: o que e como monitorar. Devido ao fato de ser impossível olhar para tudo, uma empresa precisa se concentrar em monitorar problemas específicos. Após identificar os indicadores ou medidas para as quais olhar, ela precisa desenvolver uma estratégia para monitorar esses fatores e responder efetivamente.

Nas Seções 5.7 e 5.8, examinamos detalhadamente como determinar o que deve ser medido por um sistema BPM. Por enquanto, simplesmente observamos que “o que” é comumente definido pelo FCS e as metas ou alvos definidas no processo de planejamento estratégico. Por exemplo, se um instrumento fabricado possui um objetivo estratégico especificado para aumentar a margem geral de lucro de sua linha atual de produtos em 5% anuais pelos próximos três anos, a empresa precisa monitorar a margem de lucro ao longo do ano para ver se está tendendo à taxa anual pretendida de 5%. Na mesma tendência, se essa empresa planeja introduzir um novo produto a cada semestre pelos próximos dois anos, ela precisa rastrear a introdução de novos produtos no período de tempo designado.

Sistemas de controle de diagnósticos

A maioria das empresas usa o que é conhecido como *sistema de controle de diagnóstico*, de modo a monitorar seu desempenho e corrigir os desvios dos padrões presentes de desempenho. Isso se aplica mesmo para aquelas empresas que não possuam processos de BPM formal ou sistemas. Um **sistema de controle de diagnóstico** é um sistema computacional, o que significa que tem entradas, um processo para transformar as entradas em saídas, um padrão ou marca comparativa contra a qual comparar as saídas, e um canal de retorno para permitir que informações sobre variâncias entre as saídas e o padrão a serem comunicadas e agilizadas. Quase qualquer sistema de informação pode ser usado como um sistema de controle de diagnóstico se for possível (1) definir uma meta anteriormente, (2) medir saídas, (3) computar ou calcular variâncias absolutas ou rela-

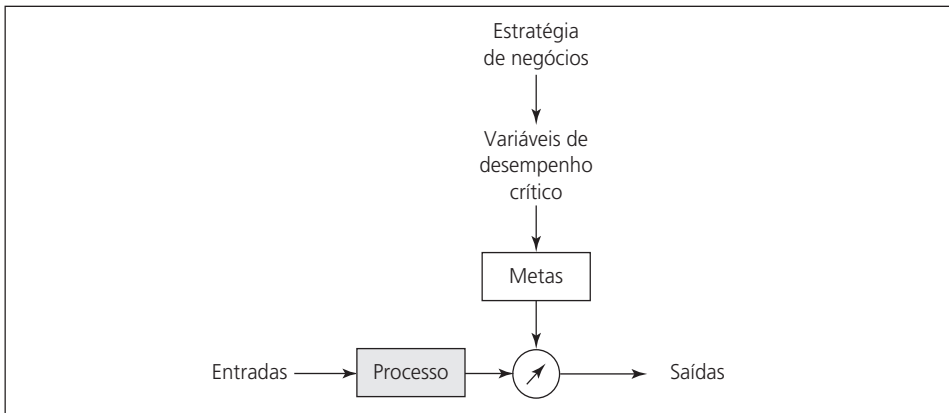


Figura 5.2 Sistema de controle de diagnóstico.

Fonte: R. Simons, *Performance Measurement and Control Systems for Implementing Strategy*, Prentice Hall, Upper Saddle River, NJ, 2002.

tivas de desempenho, e (4) usar as informações de variância como retorno para alterar entradas e/ou processos para trazer o desempenho de volta em linha com metas presentes e padrões. Os elementos principais de um sistema de controle de diagnósticos são apresentados na Figura 5.2. Placares equilibrados, painéis de desempenho, sistemas de monitoramento de projetos, sistema de recursos humanos e sistemas de relatórios financeiros são todos exemplos de sistemas que podem ser utilizados diagnosticamente.

Um sistema de controle de diagnóstico eficaz encoraja o *gerenciamento por exceção*. Em vez de monitorar constantemente uma variedade de processos internos e valores de alvo e comparar resultados reais com resultados planejados, os gerentes recebem regularmente relatórios de exceção de agenda. Medidas que estão alinhadas com expectativas recebem pouca atenção. Se, por outro lado, uma variação significativa for identificada, então – e só então – os gerentes precisam investir tempo e atenção para investigar a causa do desvio e iniciar ações corretivas apropriadas.

Armadilhas da análise de variância

Em muitas empresas, a grande maioria da análise de exceção se concentra em variâncias negativas quando grupos funcionais ou departamentos não conseguem atingir seus alvos. Há raras variâncias revisadas para oportunidades em potencial, e raramente a análise se concentra em pressuposições subjacentes a padrões de variância. Leve em consideração, por um momento, os dois caminhos da Figura 5.3. Nesta figura, a linha tracejada de A a B representa resultados planejados ou em alvo em um período de tempo especificado. Reconhecendo que haverá pequenos desvios do plano, podemos esperar que os resultados reais desviem um pouco do resultado alvo. Quando o desvio é maior do que o esperado, isso é visto como um erro operacional que necessita ser corrigido. Nesse ponto, os gerentes geralmente direcionam seus funcionários para fazer o que for possível para pôr o plano de volta na esteira. Se as receitas estão abaixo do plano, eles são repreendidos para que vendam mais. Se os custos estão acima do plano, eles devem parar de gastar.

Todavia, o que ocorre se nossas pressuposições estratégicas – não as operações – estiverem erradas? E se a empresa precisa mudar de direções estratégicas rumo ao ponto C, em vez de continuar com o plano original? Como o Caso de aplicação 5.1 exemplifica, o resultado de proceder com base em pressuposições falaciosas pode ser desastroso. A única maneira de fazer esse tipo de determinação é monitorar o desempenho mais do que real *versus* alvo. Qualquer que seja o sistema de controle de diagnóstico que esteja sendo utilizado, ele precisa rastrear pressuposições

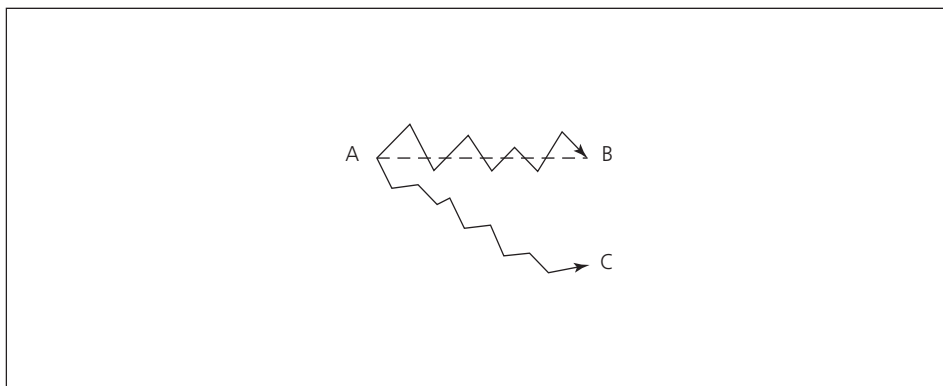


Figura 5.3 Variância operacional ou problema estratégico?

subjacentes, relações de causa e efeito e a validade geral da estratégia planejada. Leve em consideração, por exemplo, uma estratégia de crescimento que se concentra na introdução de um novo produto. Ela é comumente baseada em certas pressuposições sobre exigência de mercado e a disponibilidade de peças de fornecedores em especial. À medida que se desdobra a estratégia, o gerenciamento precisa monitorar não somente as receitas e custos associados com o novo produto, mas também as variações na demanda de mercado ou a disponibilidade de peças ou qualquer outra pressuposição em especial.

CASO DE APLICAÇÃO 5.1

Planejamento dirigido por descoberta: o caso da Euro Disney

Lembram-se do Zapmail, da FedEx? Que tal o filme instantâneo da Polaroid? Provavelmente, você não se lembra de nenhum deles, pois foram grandes falhas. Qualquer novo investimento – novos produtos, mercados, parcerias e semelhantes – tem uma grande chance de fracasso, mas muitos desses fracassos poderiam ser prevenidos com o enfoque certo de gerenciamento de planejamento, medidas e desempenho.

Há uma diferença entre planejamento para uma linha convencional de negócios e planejamento para um novo empreendimento. O planejamento convencional opera segundo a premissa de que uma empresa pode extrapolar os resultados futuros de uma plataforma bem-entendida e previsível de experiências passadas. Em planejamentos baseados em plataforma, desvios de planos são ruins, indicando que o gerenciamento necessita pôr o negócio de volta no curso. Novos empreendimentos precisam de um negócio para prever o que é desconhecido, incerto, e ainda não óbvio para a concorrência. Novos empreendimentos são feitos com uma grande razão de suposição para conhecimento sólido. Devido ao fato de

suposições associadas com um novo empreendimento geralmente são erradas, desvios de planos devem ser esperados, em vez de lamentados. Nesses casos, se o planejamento baseado em plataformas for utilizado, os resultados podem ser um tanto desastrosos, pois as suposições são tratadas como fatos, no lugar de estimativas que devem ser questionadas e testadas.

EXPERIÊNCIA DA EURO DISNEY

A Walt Disney Company é bem-conhecida por seus parques temáticos. Seu sucesso neste mercado não se restringiu aos Estados Unidos. A Disney de Tóquio é um sucesso financeiro e de relação públicas desde sua abertura em 1983. A Euro Disney (agora chamada de Disneylândia de Paris) é outra história. Em dois anos de abertura, ela alcançou seu alvo de 11 milhões de admissões. Entretanto, isso só foi realizado reduzindo os preços dos ingressos drasticamente. O gasto médio por visita foi muito abaixo do planejado. O resultado foi a perda de US\$ 1 bilhão.

Enquanto havia muito fatores subjacentes à perda, um dos principais problemas é que a Disney construiu

seus planos estratégicos e operacionais na suposição de que os visitantes da Euro Disney se comportariam essencialmente da mesma forma que os visitantes dos outros parques temáticos da empresa. Basicamente, ela tratou a Euro Disney como com um negócio em progresso, mais do que um novo empreendimento. Particularmente, supôs que os visitantes europeus pagariam por bem um alto preço de entrada (US\$ 40, na época), que ficariam uma média de 4 dias em um dos hotéis do parque, que comeriam regularmente ao longo do dia e que comprariam uma mistura semelhante de souvenirs a que os outros visitantes dos outros parques compravam. O que ocorreu foi que eles ficaram afastados do parque devido ao alto preço de entrada, os visitantes ficavam apenas uma média de dois dias porque havia muito menos passeios na Euro Disney do que nos outros parques (na época, 15 contra 45), eles geralmente saíam do parque para comer porque os restaurantes foram planejados para lidar com um fluxo contínuo de visitantes, em vez de grande multidões européias que estavam acostumadas a almoçar ao meio-dia e compravam menos souvenirs, principalmente os itens de alta margem.

PLANEJAMENTO DIRIGIDO POR DESCOBERTA

Como a Disney, outras empresas pagaram um preço semelhante por usar planejamento baseado em plataformas construídas sob suposições implícitas, que depois se tornaram problemáticas. Para evitar esses erros, McGrath e MacMillan (1995, 2000) sugeriram que as empresas utilizassem o planejamento por descoberta. O *planejamento dirigido por descoberta* oferece um modo sistemático de descobrir suposições sistemáticas que de outra forma permanecem despercebidas e não-desafiadas. É o método chamado de *dirigido por descoberta*, porque os planos que se desenvolvem

como novos dados são descobertos e novos potenciais são descobertos. O planejamento dirigido por descoberta jaz em quatro etapas inter-relacionadas:

1. *Declaração de receita reversa*. Em vez de usar receitas previstas para planejar custos e lucros, a empresa especifica suas metas de lucratividade e retira as receitas exigidas e custos permitidos que gerarão a lucratividade desejadas.
2. *Especificações de operação* proforma. A segunda etapa é distribuir todas as atividades necessárias para produzir, vender, prestar serviços e entregar o produto ou serviço ao cliente. Juntas, essas atividades representam os custos permissíveis.
3. *Lista de verificação de suposições*. Quase qualquer indústria ou mercado tem padrões e marcas de nível externos e internos que podem ser utilizados para produzir uma lista de verificação ou métrica de desempenho que pode ser rastreada e evoluída enquanto o empreendimento de desenrola. Podemos citar como exemplo as razões de ativo para vendas, giros de estoque, margens de lucro, e estratégias de vendas por dia.
4. *Planejamento de marco*. Cada plano distribui certos marcos. Com cada marco, suposições em especial da lista de verificação precisam ser testadas. Os resultados desses testes são utilizados para fazer decisões sobre investimentos contínuos de recursos.

Fontes: compilado de R. McGrath and I. MacMillan. "Discovery-Driven Planning," *Harvard Business Review*, julho-agosto de 1995; R. McGrath and I. MacMillan. *The Entrepreneurial Mindset*, Harvard Business School Press, Boston, 2000; e P. Schoemaker, *Profiting from Uncertainty*, The Free Press, New York, 2002.

Questões de revisão da Seção 5.5

1. Quais são os principais elementos de um sistema de controle de diagnóstico?
2. O que é *gerenciamento por exceção*?
3. Qual é uma das maiores armadilhas da análise de variância, de uma perspectiva gerencial?
4. Que tipos de problema de planejamento são abordados pelo planejamento dirigido por descoberta?
5. Quais são as principais etapas do planejamento dirigido por descoberta?

5.6

ALÇÃO E AJUSTE: O QUE PRECISAMOS FAZER DE FORMA DIFERENTE?

As empresas gastam uma enorme quantidade de dinheiro e tempo desenvolvendo planos, coletando dados e gerando relatórios de gerenciamento. O Hackett Group (thehackettgroup.com) gastou uma quantidade substantiva de tempo testando os hábitos de planejamento e relatório de

mais de 2.000 empresas, incluindo várias das maiores empresas dos Estados Unidos. A estrutura de seus estudos está detalhada em *Insights de tecnologia 5.2*. De acordo com os testes do Hackett Group, a empresa média de bilhões de dólares dedica 23.000 e 25.000 dias úteis a cada ano para planejamento e relatórios de gerenciamento. Essas mesmas firmas levam perto de 5 meses para criar seus planos estratégicos, 4 meses para fazer um plano tático, 5 dias para fechar seus registros financeiros no final de um período contábil, 11 dias após o final do mês para produzir relatórios de gerenciamento e 15 dias após o final do mês para produzir previsões revisadas. Parte do problema é que a empresa média está muito estimulada com detalhes, desenvolvendo planos financeiros para uma média de 370 linhas de itens (p. ex., depósitos orçamentários) e sobrecarregando os gerentes com uma média de 140 diferentes métricas de desempenho. O resultado geral é de que os membros da equipe profissional gastam metade do seu tempo coletando e validando dados e pouco tempo analisando os dados.

Insuficiência analítica

O impacto geral das práticas de planejamento e de relatório da empresa média é que o gestor tem pouco tempo para revisar os resultados de uma perspectiva estratégica, decidir o que deveria ser feito de forma diferente, e agir nos planos revisados. Como um relatório de pesquisa de Neely e Bourne (2000) descreve:

O processo completo de medir o desempenho é completamente desperdiçado, a menos que se tome uma atitude em relação aos dados de desempenho coletados. É muito frequente, em muitas empresas, que os gestores não consigam fazer isso. Eles produzem os gráficos. Eles produzem os relatórios, mas não conseguem analisar os dados e decidir o que fazer de forma diferente dentro da empresa para garantir que os números do mês seguinte serão melhores dos que os deste. É quase como se os gestores de hoje tenham se tornado tão obcecados com a medida de desempenho que eles não tem mais tempo de agir nos dados de desempenho quando forem coletados.

Números de pesquisa indicam que o gerenciamento de alto nível gasta cerca de 3 horas por mês debatendo estratégias e 80 % do seu tempo em problemas que lidam menos que 20% do valor de longo prazo da empresa. Mesmo empresas que gastam tempo revisando resultados frequentemente dão ênfase a detalhes operacionais e itens de linha de orçamento em vez de objetivos estratégicos e iniciativas conectadas. Novamente, como indica o teste do The Hackett Group, grande parte da revisão se concentra em métricas financeiras, mais do que operacionais (65 %), indicador atrasado (*lagging indicator*) no lugar de indicador líder (*leading indicator*) (75 %) e medidas internas ao invés de medidas externas (80 %). O problema é que as medidas internas de desempenho financeiro para períodos de tempo anteriores oferecem pouca prova do que deveria ser feito no futuro.

Devido ao fato de a empresa média gastar uma quantidade tão exorbitante de tempo criando seus planos estratégicos e operacionais, a última coisa que quer fazer é revisitar esses planos, exceto com base anual. Em comparação, as empresas que utilizam as melhores práticas revisitam e ajustam seus planos estratégicos e previsões com frequência ao longo do ano e reatribuem recursos com base nesses planos e previsões ajustadas. De fato, as empresas que tomam as melhores práticas usam “previsões contínuas” para fazer esses ajustes, um processo debatido no *Arquivo Online W5.2*. Mais uma vez, elas conseguem fazer esses ajustes porque se concentram em um conjunto pequeno de indicadores principais de desempenho que estão ligados aos objetivos estratégicos, em vez de um grande conjunto de métricas e itens de linha que vão às centenas. Imagine por um momento que você faz parte da equipe de gerenciamento de uma grande empresa de fabricação. Qual é a dificuldade para a equipe de gerenciamento planejar e prever demanda por um único produto? E múltiplos produtos? O que dizer sobre 100 itens de linha (p. ex., custos de fabricação, custos de

INSIGHTS DE TECNOLOGIA 5.2

Testando melhores práticas para BPM

Desde 1991, o The Hackett Group, parte da firma de consultoria Answerthink (answerthink.com), tem testado as práticas de relatórios de planejamento e gerenciamento de empresas clientes. Para cada cliente, o resultado de um estudo de *benchmark* é uma comparação abrangente dos processos do cliente relativos à organização média e as organizações de alto nível. Hoje em dia, o banco de dados de *benchmark* do The Hackett Group inclui resultados de mais de 2.400 clientes. Os clientes incluem 97% da Dow Jones Industrials, 81% da *Fortune 100* e 88% da Dow Jones Global Titans Index.

O processo de *benchmark* mede tanto a *eficiência* e a *efetividade* das práticas de relatório de planejamento e gerenciamento de empresas. Eficiência é medida examinando-se o custo e produtividade de uma empresa (p. ex., número de funcionários, dias para fechar os livros, custos de transação), e efetividade é medida olhando para sua qualidade e valor (p. ex., volume de vendas em aberto, porcentagem de erros de contas a pagar). As empresas de alto nível são aquelas cuja eficiência e efetividade estão nos primeiros 25%.

O processo de *benchmark* do The Hackett Group divide planejamento e relatório de gerenciamento em quatro subprocessos: planejamento estratégico, planejamento operacional e financeiro, relatórios e previsão. Cada subprocesso é avaliado em termos de cinco dimensões de eficiência e efetividade:

1. *Alinhamento estratégico*. Concentra-se no elo do processo para os objetivos e estratégias da empresa.
2. *Parceria*. Julga o nível de colaboração entre diferentes clientelas dentro e fora da empresa.
3. *Processo*. Mede tempo de ciclo, custo e qualidade.
4. *Tecnologia*. Mede o grau ao qual a tecnologia foi alavancada efetivamente nos processos de planejamento e relatório.
5. *Pessoas e empresas*. Lida com uma ampla rede de fatores, como níveis de preenchimento de vagas, habilidades, experiência, estrutura organizacional, liderança, educação, treinamento, papéis e responsabilidades, além de compensação.

Em geral, um *benchmark* fornece a uma empresa uma visão abrangente de seu desempenho, o que

permite ao gerenciamento da empresa identificar claramente as oportunidades para melhorar e assegurar o equilíbrio apropriado entre eficiência e efetividade. O princípio norteador de um programa de melhores práticas é primeiro eliminar todo o trabalho desnecessário, para depois padronizar e simplificar todo o trabalho restante e só depois buscar aplicar a tecnologia para alcançar o desempenho superior. A suposição é de que não há razão para simplificar ou automatizar atividades que não deveriam existir.

Nos últimos anos, os resultados de *benchmark* do The Hackett Group indicam que:

- Empresas de alto nível estão significativamente mais eficientes do que seus pares nos custos de gerenciamento. Grande parte da redução é o resultado de permuta de trabalho para tecnologia, processo de otimização e terceirização.
- Empresas de alto nível que se concentram em excelência operacional e experiência reduziram significativamente as taxas de rotação de empregados voluntários, o que se deve em parte à alocação de mais equipe para lidar com problemas de ciclo de vida de funcionários.
- Empresas de alto nível possuem estratégias de fontes híbridas que combinam serviços compartilhados e terceirização. Essas duas práticas são usadas para reduzir custos.
- Empresas de alto nível fornecem gerenciamento com as ferramentas e treinamento para alavancar as informações corporativas e para guiar o planejamento estratégico, o orçamento e a previsão.
- Empresas de alto nível alinham intimamente os planos estratégicos e táticos, permitindo que áreas funcionais (p. ex., *marketing*) contribuam mais efetivamente para metas gerais de negócios. Essas áreas funcionais são mais efetivas alinhando áreas operacionais com metas de negócios estratégicas e fornecem maior valor para suas companhias.

Resultados recentes também indicam que a lacuna entre as empresas médias e de alto nível está crescendo. Resultados detalhados do estudo de *benchmark* do The Hackett Group para planejamento e relatórios de gerenciamento podem ser encontrados em Hoblitzell (2003) e Axson (2003) e em thehackettgroup.com.

Fontes: compilado de T. Hoblitzell, *Best Practices in Planning and Budgeting*, 2003, answerthink.com/pdf/pdf_services/04_intelligence/whitepapers/wp_bpm_planbudget.pdf (acesso em março de 2006); D. Axson, *Best Practices in Planning and Management Reporting*, Wiley, Hoboken, NJ, 2003; e thehackettgroup.com.

distribuição) para esses produtos, não só receita de vendas? É uma tarefa impossível de realizar com qualquer grau de exatidão e certeza. Eis por que as empresas que utilizam as melhores práticas gastam seu tempo garantindo que os planos estratégicos sejam intimamente integrados com os planos operacionais e processos de monitoração (algo que ocorre em apenas 25 % de todas as empresas, segundo o The Hackett Group).

Questões de revisão da Seção 5.6

1. O que os *benchmarks* do The Hackett Group medem?
2. Com base nos resultados dos *benchmarks* do The Hackett Group, de que maneiras as empresas de alto nível diferem de empresas médias?
3. Segundo o The Hackett Group, quanto tempo e recursos a empresa média gasta em planejamento e relatório de gerenciamento?
4. Qual é uma das principais razões pelas quais as empresas gastam tão pouco tempo analisando os dados de gerenciamento que coletam?

5.7

MEDIDA DE DESEMPENHO

No BPM, há um sistema de medida de desempenho. Segundo Simons (2002), os **sistemas de medida de desempenho**:

Auxiliam os gerentes a rastrear as implementações de estratégia de negócios comparando os resultados reais com metas estratégicas e objetivos. Um sistema de medida de desempenho geralmente engloba métodos sistemáticos de união de metas de negócios com relatórios de retorno periódicos que indicam progresso contra metas.

Todas as medidas se tratam de comparação. Números brutos raramente são de pouco valor. Se lhe dissessem que um vendedor completou 50 % dos negócios em que ele estava trabalhando em um mês, isso teria pouco significado. Agora, suponha que lhe dissessem que o mesmo vendedor teve uma taxa de fechamento mensal de 30 % no ano passado. É óbvio que a tendência é boa. E se também lhe dissessem que a taxa média de fechamento para todos os vendedores da empresa foi de 80 %. Obviamente, esse vendedor deve pegar o ritmo. Como a definição de Simons sugere, no gerenciamento de desempenho, as principais comparações giram em torno de estratégias, metas e objetivos.

Problemas com sistemas de medida de desempenho existentes

Se você fosse pesquisar a maioria das empresas de hoje, você teria dificuldade em encontrar uma empresa que não use um sistema de medida de desempenho. O sistema mais popular em uso é uma variante do indicador balanceado (BSC) de Kaplan e Norton (1996). Por exemplo, os benchmarks do The Hackett Group (Hackett Group, 2004) indicaram que 96 % de todas as empresas implementaram ou planejavam implementar um BSC. Naturalmente, entre essas empresas, parece haver alguma confusão sobre o que constitui “balanço”. O BSC Collaborative (bscol.com) estabeleceu um conjunto de critérios para certificar as aplicações de BSC (p. ex., sistemas de *software* que fornecem capacidades de BSC). Nas suas palavras (BSC Collaborative, 2000, p. 2):

O principal da metodologia BSC é uma visão holística de um sistema de medidas ligado à direção estratégica da empresa, baseado em quatro perspectivas do mundo, com a medida financeira suportada por cliente, interno, e métricas de aprendizado e crescimento. Medindo e gerenciando o negócio usando esse conjunto holístico de métricas, uma empresa pode assegurar implementação rápida e eficaz da estratégia e facilitar o alinhamento organizacional e a comunicação.

Todavia, entre as empresas do estudo de *benchmark* do The Hackett Group, a grande maioria das medidas são financeiras ou operacionais em natureza (melhor que 80%). O que essas empresas realmente têm é um “placar” – um conjunto de relatórios, gráficos e expositores especializados que os permitem comparar resultados reais com resultados planejados para uma coleção mista de medidas.

Relatórios financeiros dirigidos por calendário são um importante componente da maioria dos sistemas de medida de desempenho. Não é surpresa. Primeiramente, a maioria desses sistemas está sob a competência do departamento financeiro. Em segundo lugar, a maioria dos executivos tem pouca fé em alguma coisa, exceto em números financeiros ou operacionais. Pesquisa indicam que os executivos valorizam uma variedade de diferentes tipos de informação (p.ex., financeira, operacional, mercado, cliente), mas pensam que fora das arenas financeiras ou operacionais a maioria dos dados são suspeitos, e eles não querem apostar seus empregos na qualidade dessas informações (Schiemann e Lingle, 1999).

Os empecilhos de se usar dados financeiros como o centro de um sistema de medida de desempenho são bem conhecidos. Dentre as limitações citadas com mais frequência, temos:

- As medidas financeiras são comumente relatadas por estruturas organizacionais (p. ex., gastos com pesquisa e desenvolvimento) e não pelo processo que as produziu.
- As medidas financeiras são indicador atrasado (*lagging indicator*), dizendo-nos o que aconteceu, não por que aconteceu ou o que deve acontecer no futuro.
- Medidas financeiras (p. ex., despesas gerais administrativas) são geralmente o produto de alocações que não são relacionadas ao processo subjacente que as gerou.
- Medidas financeiras se concentram no curto prazo e fornecem pouca informação sobre o prazo mais longo.

A miopia financeira não é o único problema que enfeza muitos dos sistemas de medida de desempenho em operação hoje em dia. A sobrecarga de medida e a obliquidade de medida são também grandes problemas que confrontam a atual safra de sistemas.

É muito incomum encontrar empresas anunciando orgulhosamente que estão rastreando 200 ou mais medidas no nível corporativo. É difícil imaginar tentar dirigir um carro com 200 mostradores no painel. Entretanto, parece que temos poucos problemas dirigindo empresas com 200 mostradores no painel corporativo, mesmo que saibamos que os humanos têm grandes dificuldades em lidar com mais de uma certa quantidade de problemas e que qualquer outra coisa é simplesmente posta de lado. Este tipo de sobrecarga é exacerbado pelo fato de que as empresas raramente excluem as medidas que coletam. Se algum novo dado ou solicitação por dados aparece, é simplesmente adicionado à lista. Se o número de medidas é 200 hoje, será 201 amanhã e 202 depois. Mesmo que os planos mudem e oportunidades e problemas apareçam e desapareçam com frequência cada vez maior, faz-se pouco esforço para determinar se a lista de medidas que estamos rastreando ainda é aplicável à situação atual.

Para muitas das medidas sendo rastreadas, o gerenciamento tem falta de controle direto. Michael Hammer (2003) chamou isso de *princípio de obliquidade*. Por um lado, medidas como rendimentos por ação, retorno sobre patrimônio, lucratividade, participação no mercado e satisfação do cliente precisam ser monitoradas. Por outro lado, essas medidas somente podem ser buscadas de modo oblíquo. O que pode ser controlado são as ações de trabalhadores individuais ou funcionários. Infelizmente, o impacto de qualquer ação individual em uma estratégia corporativa ou estratégia de unidade de negócios é negligenciável. O que é necessário para unir o crítico com o controlável é um modelo de negócios estratégico ou metodologia que se inicia no topo e liga metas corporativas e objetivos até as iniciativas de baixo sendo conduzidas por indicadores individuais.

Medida de desempenho efetiva

Muito livros oferecem receitas para determinar se um grupo de medidas de desempenho é bom ou ruim. Entre os ingredientes básicos de um bom grupo estão os seguintes:

- As medidas devem se concentrar em fatores cruciais.
- As medidas devem ser uma mistura de passado, presente e futuro.
- As medidas devem equilibrar as necessidades dos acionistas, funcionários, parceiros, fornecedores e outras partes interessadas.
- As medidas deveriam se iniciar no topo e fluir para baixo.
- As medidas precisam ter metas que se baseiem em pesquisa e realidade, ao invés de serem arbitrárias.

O Arquivo Online W5.3 oferece uma metodologia mais detalhada e conjunto de modelos para fazer a determinação.

Enquanto todas as características recentemente listadas são importantes, a verdadeira chave para o sistema de medida de desempenho efetivo é ter uma boa estratégia. As medidas precisam ser derivadas das estratégias corporativas e de unidade de negócios e de uma análise dos principais processos de negócios necessários para alcançar essas estratégias. Naturalmente, é mais fácil falar do que fazer. Se fosse simples, a maioria das empresas já teria sistemas eficientes de medida de desempenho implementados, mas não tem.

O Caso de Aplicação 5.3, que descreve o portal dos indicadores chave de desempenho (KPI) baseados na Web da International Truck and Engine Corporation, oferece percepções sobre as etapas que uma empresa segue para implementar seu sistema de medida de desempenho. Também oferece uma lista dos fatores que são essenciais para o sucesso de um sistema de medida de desempenho.

CASO DE APLICAÇÃO 5.3

International Truck and Engine Corporation

A International Truck and Engine Corporation é uma empresa em operação da Navistar International Corporation. A empresa produz caminhões comerciais, ônibus escolares, e uma variedade de motores a diesel para caminhões, vans, e veículos esporte. A empresa possui aproximadamente 15.000 funcionários e opera instalações de fabricação na América do Norte e no México.

PORTAL DOS KPI

No passado, os gerentes da International usavam uma miscelânea de relatórios operacionais baseados em papel para monitorar e gerenciar seu desempenho operacional. Há alguns anos atrás, a companhia lançou um portal baseado na Web de KPI para tentar ajudar 500 de seus gerentes e executivos a rastrear e gerenciar o desempenho da empresa em 130 KPIs. O portal dos KPIs oferece aos executivos e gerentes de

grupos de negócios visões personalizadas dos principais impulsionadores dos negócios. Dessa forma, eles podem chegar à causa essencial de um problema em horas ou dias, ao invés de semanas ou meses.

O projeto do portal dos KPIs foi iniciado pelo vice-presidente e controlador da International. Sua meta era acelerar o ciclo de fechamento financeiro mensal da empresa, um processo que estava tomando até duas semanas no final de cada mês. Dados mais em tempo ofereceria aos gerentes da International uma melhor visão da lucratividade da empresa e os fatores que afetavam a lucratividade. Ele disse que a concentração era capturar e disseminar métricas principais de negócios com mais frequência do que os tradicionais relatórios mensais para os gerentes de linha de frente, que tomavam decisões que afetavam as estratégias, planos e previsões gerais de negócios da empresa.

DETERMINANDO REQUISITOS

Como uma primeira etapa, a International fez uma sondagem entre seus gerentes financeiros para determinar as informações e métricas necessárias para gerenciar operações e acelerar o fechamento financeiro. Uma equipe de oito membros internos e terceiros foi estabelecida para coletar e analisar as informações. A sondagem identificou 133 KPIs distintos. Para garantir a relevância contínua do portal, a equipe conduz a sondagem anualmente para determinar quais medidas deveriam ser adicionadas, continuadas, ou descartadas. Além disso, a equipe penetrou no grupo de KPIs para determinar os dados detalhados subjacentes às métricas de alto nível. Um botão Detalhes foi adicionado ao mostrador para permitir que os usuários finais visualizassem os detalhes subjacentes.

LANÇANDO O SISTEMA

Em vez de lançar o sistema em um departamento ou função por vez, a equipe decidiu entregar um KPI por vez. Começou pela demonstração da receita do grupo de Caminhões e dados de contagem de unidades. Nesse caso, os dados necessários já estavam no *data warehouse* da empresa. Medidas subsequentes exigiram que a equipe carregasse os dados de vários sistemas de origem para o *data warehouse*.

O portal é atualizado diariamente por 32 sistemas dentro do *data warehouse*, que fica em uma base de dados Informix. Dados resumidos são adicionados do *data warehouse* em um *data mart* da Hyperion Solutions (veja *Insights de tecnologia 5.5*, mais adiante, neste capítulo). Os usuários finais acessam o *data mart* através da Web. O portal dos KPIs fornece aos usuários finais gráficos para todas as métricas, bem como relatórios e outros documentos relevantes. Os usuários podem personalizar o portal para mostrar apenas as métricas que lhe interessam. Os usuários

também têm a capacidade de se exercitar pelas informações de modo a acessar os detalhes subjacentes.

O uso geral foi determinante para o sucesso do programa. Para ir ao encontro desse problema, a equipe assegurou que a interface baseada na Web fosse intuitiva de usar e fornecesse treinamento aos usuários acerca da natureza e origem dos dados subjacentes. Para ajudar a criar confiança no sistema, a equipe reconciliou os dados e informações no novo sistema com os dados em sistemas existentes. Com o tempo, os gerentes e executivos começaram a reconhecer o valor do portal e se convenceram de que o portal fornecia uma versão única da verdade.

FATORES DE SUCESSO

O sistema de KPI tem tido sucesso. Os principais fatores responsáveis por seu sucesso incluem os seguintes:

- Um patrocinador executivo altamente comprometido
- Fundos adequados
- Um desenvolvimento incremental e agenda de lançamentos
- Uma avaliação de exigências completamente baseada na Web
- Uma equipe conhecedora e talentosa
- Um núcleo de especialistas nos assuntos
- Um *data warehouse* existente
- Um piloto para demonstrar o valor e validade do produto
- Confiança nos dados
- Comunicação para explicar os benefícios, funcionalidade, e lançamento do novo sistema

Fontes: compilado de W. Eckerson, “BI Case Study: International Truck and Engine Corporation,” *Business Intelligence Journal*, inverno de 2004; e W. Eckerson, *Performance Dashboards*, Wiley, Hoboken, NJ, 2006.

Questões de revisão da Seção 5.7

1. O que é um sistema de medida de desempenho?
2. Quais são alguns dos empecilhos de confiar unicamente em métricas financeiras para medir o desempenho?
3. Qual o princípio da obliquidade?
4. Quais são algumas das características de um “bom” conjunto de medidas de desempenho?
5. Que etapas a International Truck and Engine utilizou para criar as medidas para o seu sistema? Que fatores foram essenciais para o seu sucesso?

METODOLOGIAS DE BPM

Há mais coisas em relação à medida de desempenho do que simplesmente manter o escore. Um sistema de medida de desempenho efetivo deveria ajudar a fazer o seguinte:

- Alinhar objetivos de nível superior com iniciativas básicas.
- Identificar oportunidades e problemas em tempo.
- Determinar prioridades e alocar recursos com base nas prioridades.
- Mudar medidas quando os processos e estratégias subjacentes mudam.
- Delinear responsabilidades, entender o desempenho real relativo às responsabilidades, e recompensar e reconhecer realizações.
- Agir para melhorar os processos e procedimentos quando os dados os autorizam.
- Planejar e prever de forma mais confiável e oportuna.

É necessária uma estrutura de medida de desempenho holística ou sistemática para realizar essas metas, bem como outras. Nos últimos 40 anos, vários sistemas foram propostos. Alguns deles, como custeio baseado em atividades ou gerenciamento, têm foco financeiro. Outras, como o gerenciamento de qualidade total (TQM), são orientadas para processos. No debate que segue, examinamos dois enfoques amplamente utilizados que dão apoio aos processos básicos subjacentes ao BPM: BSC (veja bsc.com) e Six Sigma (veja mu.motorola.com).

Balanced scorecard

Provavelmente, o sistema de gerenciamento de desempenho mais bem conhecido e amplamente utilizado é o BSC. Kaplan e Norton articularam a metodologia pela primeira vez no seu artigo “*The Balanced Scorecard: Measures That Drive Performance*” (O *balanced scorecard*: medidas que impulsionam o desempenho), no *Harvard Business Review*, que surgiu em 1992. Poucos anos depois, em 1996, esses mesmos autores produziram um livro revolucionário – *The Balanced Scorecard: Translating Strategy into Action* (O *balanced scorecard*: a estratégia em ação) —, que documentava como as empresas estavam utilizando o BSC para não apenas suplementar suas medidas financeiras com medidas não-financeiras, mas também comunicar e implementar suas estratégias. Nos últimos anos que se passaram, o BSC se tornou um termo genérico (como Coca ou Xerox) que é utilizado para representar quase todo o tipo de aplicação e implementação de painel. Kaplan e Norton lançaram um novo livro em 2000, *The Strategy-Focused Organization: How Balanced Scorecard Companies Thrive in the New Business Environment* que foi planejado para reenfatar a natureza estratégica da metodologia BSC. Este livro foi seguido alguns anos mais tarde, em 2004, por *Strategy Maps: Converting Intangible Assets into Tangible Outcomes* (Mapas estratégicos: convertendo ativos intangíveis em retornos tangíveis), que descreve um processo detalhado para ligar objetivos estratégicos a táticas e iniciativas operacionais.

O significado de equilíbrio

De um ponto de vista de alto nível, **balanced scorecard (BSC)** é tanto uma medida de desempenho e metodologia de gerenciamento que ajuda a traduzir os objetivos e metas financeiras, de clientes, de processos internos e de aprendizado e crescimento de uma empresa em um conjunto de iniciativas acionáveis. Como uma metodologia de medida, o BSC é planejado para superar as limitações de sistemas que têm foco financeiro. Ele faz isso traduzindo a visão e estratégia de uma empresa em um conjunto de objetivos financeiros e não-financeiros, medidas, metas e iniciativas. As relações entre os objetivos financeiros e não-financeiros são mostradas na Figura 5.4. Os objetivos não-financeiros caem em uma das três perspectivas:

- *Cliente*. Os objetivos definem como a empresa deveria aparecer para os seus clientes se for realizar sua visão

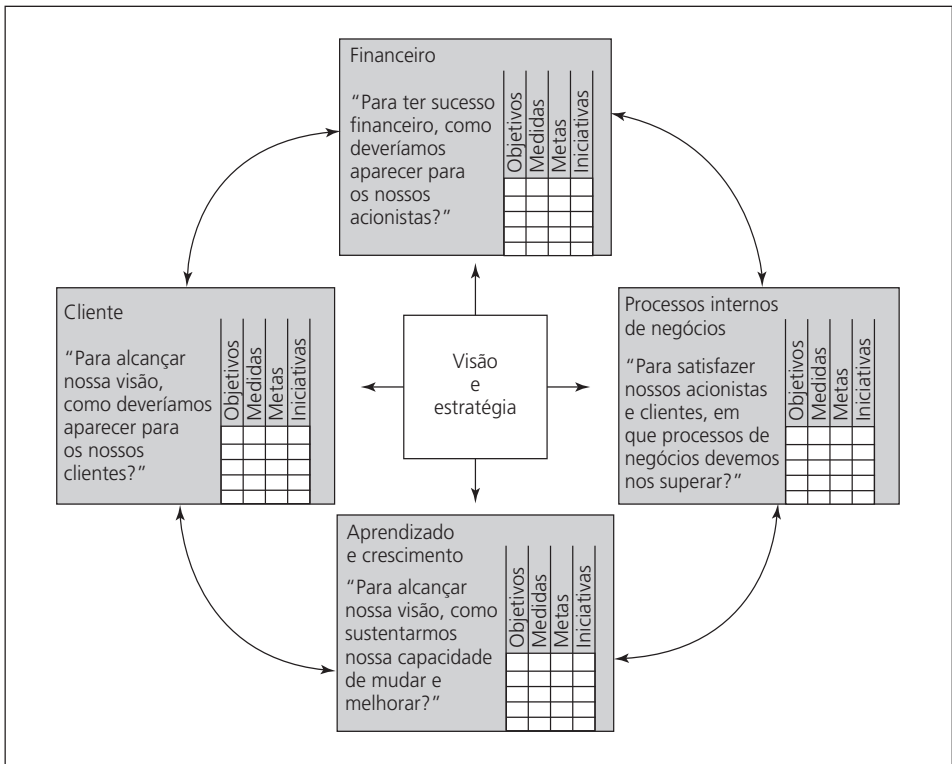


Figura 5.4 Perspectivas do *balanced scorecard*.

Fonte: Balanced Scoreboard Institute. balancedscorecard.org/basics/bsci.html (acessado em abril de 2006)

- *Processos internos de negócios*. Esses objetivos especificam os processos nos quais a empresa deve se superar de modo a satisfazer seus clientes e acionistas
- *Aprendizado e crescimento*. Esses objetivos indicam como uma empresa pode melhorar sua capacidade de mudar e melhorar para alcançar sua visão

No BSC, o termo *balance* (equilíbrio) surge porque o conjunto combinado de medidas deve englobar indicadores que são:

- Financeiros e não-financeiros
- Líderes e de ocorrência (*leading e lagging*)
- Internos e externos
- Quantitativos e qualitativos
- De curto e longo prazo

Alinhando estratégias e ações

Como uma metodologia de estratégia de gerenciamento, o BSC possibilita que uma empresa alinhe suas ações com suas estratégias gerais. O BSC realiza essa tarefa através de uma série de etapas inter-relacionadas. As etapas específicas que estão envolvidas variam de livro para livro. Em nosso caso, o processo pode ser apreendido em cinco etapas:

1. Identificar objetivos estratégicos para cada uma das perspectivas (cerca de 15 a 25 no total).
2. Associar medidas com cada um dos objetivos estratégicos; uma mistura de quantitativo e qualitativo deve ser utilizada.

- 3. Atribuir metas para a medida.
- 4. Listar iniciativas estratégicas para realizar cada um dos objetivos (i.e., responsabilidades).
- 5. Ligar os vários objetivos estratégicos através de um diagrama de causa e efeito chamado de *mapa estratégico*.

Como um exemplo do processo, considere o mapa estratégico mostrado na Figura 5.5. Um **mapa estratégico** delinea as relações entre os principais objetivos organizacionais para todas as quatro perspectivas de BSC. Nesse exemplo, o mapa especifica as relações entre sete objetivos que cobrem quatro diferentes perspectivas. Como outros mapas estratégicos, este começa no topo com um objetivo financeiro (i.e., excesso de crescimento em segmentos-chave). Esse objetivo é impulsionado por um objetivo de cliente (i.e, construir fortes relações com o cliente). Por sua vez, o objetivo do cliente é o resultado de um objetivo (i.e., de processo) interno (i.e., identificar/capturar novas oportunidades de negócios). O mapa continua até o fim de uma hierarquia, onde os objetivos de aprendizado são encontrados (p. ex., desenvolver habilidades-chave).

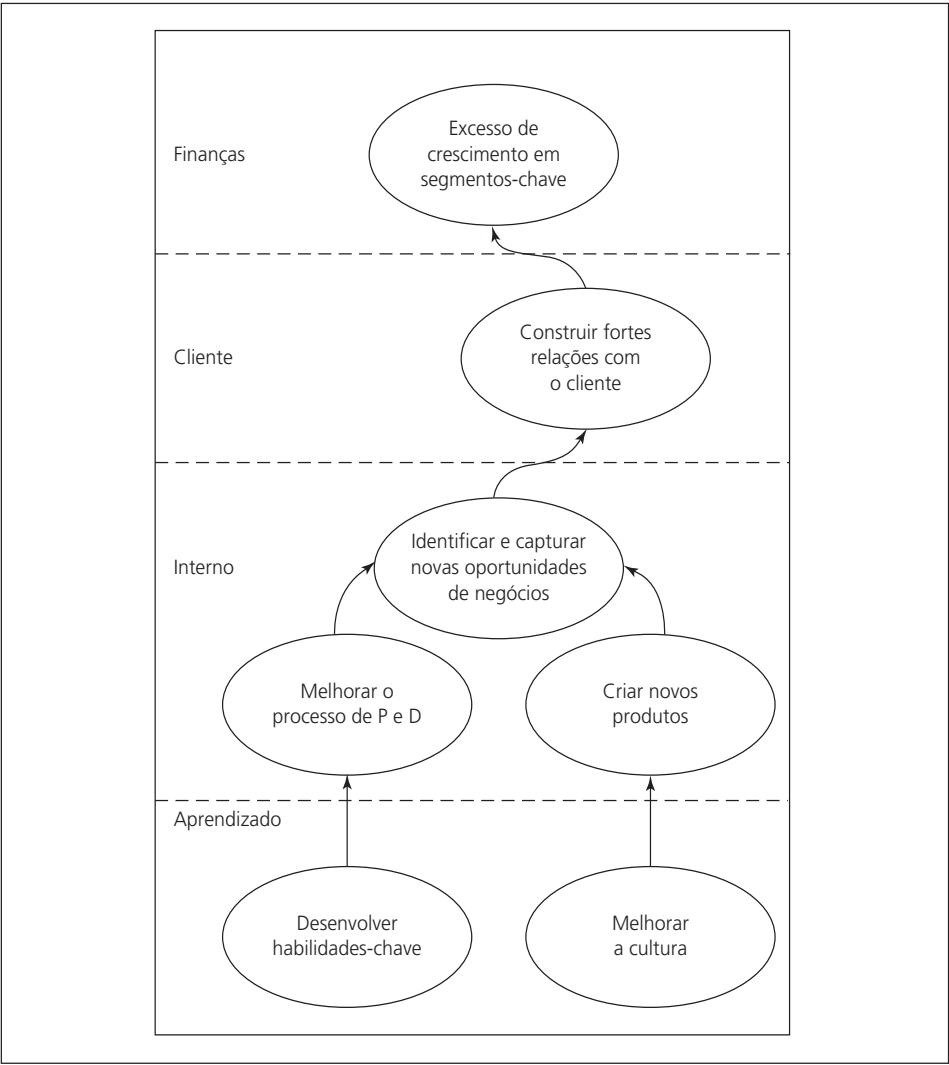


Figura 5.5 Mapa de amostra de estratégia.

Tabela 5.2 Especificação de objetivos, medidas, metas e iniciativas

<i>Objetivo</i>	<i>Medida</i>	<i>Meta</i>	<i>Iniciativa</i>
O que a estratégia está tentando alcançar	Como o sucesso ou a falha contra os objetivos é monitorado	Nível de desempenho ou taxa de melhoria necessária	Principais programas de ação necessários para alcançar as metas
Construir fortes relações com o cliente	Satisfação do cliente	Índice de serviço do cliente	Implementar a base de dados de retorno ao cliente

Cada objetivo que aparece num mapa estratégico tem uma medida, meta e iniciativa associada. Por exemplo, o objetivo “construir fortes relações com o cliente” poderia ser medido por satisfação do cliente. Para essa medida, podemos objetivar uma melhoria de 15% sobre o número do ano passado em nosso índice de serviços ao cliente. Uma das maneiras de se realizar essa melhoria é implementando a base de dados de retorno ao cliente. Os vários componentes deste exemplo são mostrados na Tabela 5.2.

No geral, os mapas estratégicos, como o da Figura 5.5, representam um modelo hipotético de um segmento do negócio. Quando nomes específicos (de pessoas ou equipes) são atribuídos às várias iniciativas, o modelo serve para alinhar as ações básicas da empresa com os objetivos estratégicos de alto nível. Quando resultados reais são comparados com resultados dos objetivos, pode-se fazer uma determinação sobre se a estratégia que a hipótese representa deveria ser questionada ou se as ações daqueles que são responsáveis por várias partes da hipótese devem ser ajustadas.

Certificação BSC

A BSC Collaborative oferece aos fornecedores de *software* a oportunidade de ter seus aplicativos certificados contra um conjunto bem-definido de critérios. Para se qualificar, um aplicativo deve suportar os elementos principais debatidos na seção anterior. Ou seja, o aplicativo deve oferecer a um usuário final a capacidade de visualizar (1) objetivos estratégicos das quatro perspectivas; (2) as medidas, metas e iniciativas associadas a cada objetivo e (3) a relação causa e efeito entre os objetivos. A BSC Collaborative também estabeleceu um novo esquema de Linguagem de Marcação Extensível (XML) que define uma estrutura de dados subjacente para todos os objetos de dados em um aplicativo de BSC. O uso desse padrão facilitará a integração das informações do BSC com outros aplicativos e entre empresas.

Six Sigma

Desde seu princípio na metade dos anos 80, o Six Sigma desfrutou de ampla adoção pelas empresas ao redor do mundo. No geral, não tem sido utilizado como uma metodologia de gerenciamento de desempenho. Em vez disso, a maioria das empresas a utilizam como uma metodologia de melhoria de processos, que permite o escrutínio dos seus processos, apontar problemas e aplicar soluções. Em anos recentes, algumas empresas, como a Motorola, reconheceram o valor de utilizar o Six Sigma para propósitos estratégicos. Nesses exemplos, o Six Sigma fornece os meios de medir e monitorar processos-chave relacionados com a lucratividade de uma empresa e para acelerar a melhoria no desempenho geral dos negócios. Devido à sua concentração nos processos de negócios, o Six Sigma também oferece uma maneira direta para lidar com problemas de desempenho após serem identificados ou detectados.

Six Sigma Definido

A história do Six Sigma data do final dos anos 70, embora muitas de suas idéias possam ser rastreadas a iniciativas de qualidade anteriores (veja Isixsigma, 2006). O termo *Six Sigma* foi cunhado por Bill Smith, um engenheiro da Motorola. Na verdade, Six Sigma é uma marca registrada mundialmente da Motorola. No final dos anos 70 e no início e meio dos 80, a Motorola foi impelida para o Six Sigma por pressões internas e externas. Externamente, estava sendo derrotada no mercado por concorrentes que eram capazes de produzir produtos de mais alta qualidade por um preço mais baixo. Internamente, quando uma firma japonesa tomou posse de uma fábrica da Motorola dos EUA que fabricava televisores Quasar e pôde produzir televisores com 1/20 do número de defeitos sob procedimentos regulares de operação, os executivos da Motorola tiveram de admitir que a sua qualidade não era boa. Em resposta a essas pressões, o CEO da Motorola, Bob Galvin, levou a empresa a um caminho de qualidade chamado Six Sigma. Desde aquele tempo, centenas de empresas ao redor do mundo, inclusive General Electric, Allied Signal, DuPont, Ford, Merrill Lynch e Toshiba utilizaram o Six Sigma para gerar bilhões de dólares em melhorias de crescimento de topo de linha e ganhos básicos.

No Six Sigma, um negócio é visto como uma série de processos. Um *processo de negócios* é um conjunto de atividades que transforma um conjunto de entradas, o que inclui fornecedores, ativos, recursos (p. ex., capital, material, pessoas), e informações para um conjunto de saídas (p. ex., bens ou serviços) para outra pessoa ou processo. A Tabela 5.3 lista algumas categorias de processos de negócios que podem afetar o desempenho geral corporativo.

Sigma, σ , é uma letra do alfabeto grego que os estatísticos usam para medir a variabilidade de um processo. Na arena da qualidade, *variabilidade* é sinônimo de número de defeitos. Em geral, as empresas aceitaram uma grande quantidade de variabilidade nos seus processos de negócios. Em termos numéricos, a norma tem sido 6.200 a 67.000 defeitos por milhão de oportunidades (DPMO). Por exemplo, se uma empresa de seguros lida com 1 milhão de reclamações, então, sob procedimentos normais de operação, 6.200 a 67.000 dessas reclamações seriam defeituosas (p. ex., malfeitas, ter erros nos formulários). O nível de variabilidade representa um nível de sigma de 3 a 4 de desempenho. Para alcançar um nível de desempenho do Six Sigma, a empresa teria de reduzir o número de defeitos para não mais que 3,4 DPMO. Logo, o **Six Sigma** é uma metodologia

Tabela 5.3 Categorias de processos de negócios

Contabilidade e medidas
Gerenciamento administrativo e de instalações
Auditorias e melhorias
Planejamento e execução de negócios
Políticas e procedimentos de negócios
Marketing global e vendas
Gerenciamento e análise de informações
Liderança e lucratividade
Aprendizado e inovação
Manutenção e colaboração
Parceria e alianças
Produção e serviços
Compras e gerenciamento de cadeia de fornecimento
Recrutamento e desenvolvimento
Pesquisa e desenvolvimento

de gerenciamento de desempenho com vistas a reduzir os números de defeitos em um processo de negócios para o mais próximo possível de 0 DPMO.

O modelo de desempenho DMAIC

O Six Sigma baseia-se em um modelo de melhoria de desempenho simples conhecido como DMAIC. Como o BPM, o **DMAIC** é um modelo de melhoria de negócios de circuito fechado e engloba as etapas de definição, medida, análise, melhoria e controle de um processo. As etapas podem ser descritas como segue:

1. *Definir*. Definir as metas, objetivos e limites da atividade de melhoria. No nível mais alto, as metas são os objetivos estratégicos da empresa. Em níveis mais baixos – níveis de departamento ou projeto –, as metas são concentradas em processos operacionais específicos.
2. *Medir*. Mede o sistema existente. Estabelecer medidas quantitativas que gerarão dados estatísticos válidos. Os dados podem ser utilizados para monitorar o progresso rumo às metas definidas na etapa anterior.
3. *Analisar*. Analisar o sistema para identificar maneiras de eliminar a lacuna entre o desempenho atual do sistema ou processo e a meta desejada.
4. *Melhorar*. Dar início a ações para eliminar a lacuna encontrando maneiras de fazer as coisas de forma melhor, mais barata, ou mais rápida. Utilizar o gerenciamento de projetos e outras ferramentas de planejamento para implementar o novo enfoque.
5. *Controlar*. Institucionalizar o sistema melhorado pela modificação dos sistemas de compensação e incentivos, políticas, procedimentos, planejamento de recursos de manufatura (MRP), orçamentos, instruções operacionais, ou outros sistemas de gerenciamento.

Para novos processos, o modelo que é utilizado é chamado de *DMADV* (definir, medir, analisar, projetar e verificar). Tradicionalmente, o DMAIC e o DMADV têm sido utilizados principalmente com problemas operacionais. Todavia, nada impede a aplicação dessas metodologias a problemas estratégicos, como a lucratividade da empresa.

Limitações do Six Sigma

Os especialistas e autoridades em Six Sigma são rápidos para elogiar a metodologia e apontar para empresas como a General Electric como prova de seu valor. O que eles geralmente não conseguem apontar é que muitas empresas tentaram a metodologia e falharam. Por exemplo, em uma pesquisa de novembro de 2004 de 276 executivos sênior conduzida pela Celerant Consulting (**celerantconsulting.com**), 43% dos respondentes notaram que as iniciativas de aumento de desempenho adotadas por suas empresas nos três anos que se passaram não alcançaram seus objetivos (veja Eagle, 2005). Segundo a pesquisa, a iniciativa de menos sucesso foi o Six Sigma. Todas as outras iniciativas – incluindo otimização de processos de negócios, terceirização e manufatura enxuta – alcançaram resultados melhores. A razão principal para a falha foi a falta de integração entre os vários projetos do Six Sigma pela empresa. Outra razão foi a de que as empresas não conseguiram instituir os papéis necessários para suportar a metodologia (veja o Arquivo *Online* W5.4 para um debate sobre esses papéis). O impacto geral foi de que os efeitos dos vários projetos do Six Sigma dentro de uma empresa eram localizados, tendo pouco impacto geral tanto na base como no topo da empresa.

Em um esforço de superar essas e outras limitações, algumas empresas, como a Motorola e o Duke University Hospital, combinaram suas iniciativas do Six Sigma com as iniciativas BSC. Dessa forma, suas iniciativas de qualidade são ligadas diretamente aos seus objetivos e metas estratégicas. Na mesma esteira, Gupta (2004) desenvolveu uma metodologia híbrida chamada *Six Sigma Business Scorecard* (Placar de Negócios Six Sigma), que une diretamente os aspectos

BSC encontra o Six Sigma

O BSC e o Six Sigma são ferramentas e metodologias de gerenciamento de desempenho amplamente adotadas. Entretanto, a maioria das empresas usuárias, as tratam como iniciativas separadas. Todavia, segundo Stan Elbaum, vice-presidente sênior de pesquisa do Aberdeen Group, em Boston, esses são programas complementares. Os verdadeiros benefícios de cada uma não podem ser alcançados, a menos que ambas sejam integradas. O enfoque do BSC permite que as empresas identifiquem, com rapidez e precisão, fraquezas importantes de desempenho, e descubram oportunidades para melhorias e crescimento. O que o BSC tem dificuldades de fazer é mostrar como consertar os problemas de desempenho. Por outro lado, os projetos do Six Sigma geralmente falseiam, pois as equipes de projeto “oscilam por toda a empresa em busca de fraquezas de desempenho ou concentrando atenção em áreas onde as melhorias só produzirão resultados marginais” (Leahy, 2005, p. 48). Tomados juntamente, entretanto, BSC e Six Sigma permitem que uma empresa chegue às causas subjacentes de uma incapacidade de desempenho.

Uma pesquisa recente (Docherty, 2005) de empresas que adotaram os programas BSC ou Six Sigma revelou que quase metade dos programas não obtiveram sucesso mesmo nos primeiros três anos de adoção, mas aqueles que os fizeram operar atingiram benefícios financeiros substanciais. As empresas com os maiores retornos líquidos foram aquelas que encontraram uma maneira de integrar ambos. A integração foi alcançada fazendo o seguinte:

- *Conversão de suas estratégias em objetivos quantificáveis.* Isto foi feito pelo mapeamento da estratégia e uso de um placar para monitorar as métricas associadas.
- *Colocação de degraus nos objetivos da empresa.* Eles decompuseram os objetivos da empresa em objetivos operacionais de nível mais baixo, aplicando o raciocínio causal que subjaz ao Six Sigma.
- *Definição de alvos com base na voz do cliente.* Eles utilizaram o BSC e o Six Sigma juntos para garantir que os objetivos operacionais teriam impacto direto nas expectativas dos clientes.
- *Implementação de projetos estratégicos usando o Six Sigma.* Eles utilizaram o Six Sigma para impelir melhorias na qualidade do produto e do processo.
- *Execução de forma consistente para gerar resultados de negócios.* Eles enxergavam a empresa

a partir de uma perspectiva de processo. O Six Sigma foi utilizado para controlar a variação dos processos, e medidas de processos foram incluídas no seu BSC.

Enquanto a integração das duas metodologias pode gerar benefícios substanciais, uma empresa enfrenta desafios substanciais ao adotar esse enfoque. De um ponto de vista organizacional, a empresa deve garantir o seguinte:

- Que as ações de indivíduos na empresa estejam alinhadas com as medidas que fluem dos objetivos da empresa.
- Que a cultura de gerenciamento que prevalece esteja alinhada com o enfoque baseado em fatos do BSC e do Six Sigma, ao invés de ter como recurso o enfoque “conserte agora”.
- Que a infra-estrutura de tecnologia de informação (TI) suporte a integração de BSC e Six Sigma.

Para alcançar essa última exigência, a infra-estrutura de TI deve fazer o seguinte:

- Permitir que os executivos criem e compartilhem estratégias em pedaços ou segmentos que tenham sentido no nível operacional.
- Permitir que as iniciativas estejam atreladas aos objetivos estratégicos ou KPI.
- Fornecer capacidades de gerenciamento de programas e de projetos, de modo que as equipes dos projetos possam impelir iniciativas para suas conclusões.
- Permitir que os processos sejam medidos e automatizados em tempo real.
- Dar suporte ao planejamento estratégico completo e ciclo de execução.

A primeira e a segunda exigências têm o suporte do BSC. As duas exigências seguintes são fornecidas pelas ferramentas e metodologias do Six Sigma. A última dessas exigências é fornecida por um sistema completo de BPM.

As empresas que combinaram com sucesso as duas metodologias dizem que não conseguem entender por que uma empresa iria querer usar uma sem a outra. Entretanto, elas também aconselham que leva cerca de um ano para fornecer o treinamento da força de trabalho necessário e superar barreiras culturais e organizacionais existentes.

de melhoria de processos do Six Sigma para a perspectiva financeira do BSC. Os benefícios e a estrutura da combinação são debatidos em *Insights de tecnologia 5.4*.

Questões de revisão da Seção 5.8

1. Quais são as características de um sistema eficaz de medida de desempenho?
2. Quais são as quatro perspectivas no BSC?
3. A que o termo *equilibrado (balanced)* se refere no BSC?
4. Que etapas são utilizadas na criação de um BSC?
5. O que é o mapa estratégico?
6. A que se refere Six Sigma?
7. Quais são os processos básicos no modelo DMAIC?
8. Como o BSC e o Six Sigma podem ser integrados? Quais são algumas das barreiras de se integrar ambos?

5.9

ARQUITETURA E APLICAÇÕES DE BPM

Não há uma única tecnologia ou aplicação que constitua o BPM. Na verdade, o BPM é suportado por diversas tecnologias e aplicações. Na falta de uma lista ou coleção definitiva, o BPM Standards Group (bpmstandardsgroup.org) propôs uma arquitetura da tecnologia de BPM (consulte a Figura 5.6). Essa arquitetura destaca as tecnologias capacitadoras e algumas das aplicações cruciais necessárias para oferecer suporte aos processos de BPM de ciclo fechado que conectam a estratégia à execução. Como indica a Figura 5.6, a arquitetura proposta é constituída por diversos níveis e consiste em uma camada de interface de usuário, uma de banco de dados e uma de dados-fonte. Não existe uma única tecnologia ou aplicação que ofereça uma cobertura completa a todas as camadas e componentes resumidos na Figura 5.6. Entretanto, alguns fornecedores de *software* de BPM proporcionam suítes abrangentes de aplicações que cobrem amplos aspectos da arquitetura proposta. Nas próximas seções, discutiremos esta arquitetura e também revisaremos algumas das suítes de aplicação oferecidas pelos fornecedores de BPM.

Arquitetura de BPM

O termo **arquitetura de sistema** refere-se aos projetos lógico e físico de um sistema. O projeto lógico detalha os elementos funcionais de um sistema e suas interações. O projeto físico especifica como o projeto lógico será de fato implementado e implantado em um conjunto específico de tecnologias, como computadores, servidores, bancos de dados, protocolos de comunicação, etc. A arquitetura de BPM proposta pelo BPM Standards Group é uma arquitetura lógica que pode ser implantada fisicamente de diversas maneiras.

Pode-se pensar no sistema de BPM e sua arquitetura como um veículo (veja COVENEY et al.: 2003). Uma das principais funções de um veículo é transportar passageiros de sua posição atual para um local desejado. Um veículo é uma integração de componentes individuais, cada um dos quais com sua devida importância, porém relativamente inúteis caso não estejam integrados como um pacote completo. Contanto que o condutor saiba como operar o veículo e aonde ir, há três grandes componentes no veículo que contribuem para o sucesso da viagem: o chassi, ao qual todos os elementos são afixados; o motor, que move o veículo para frente; e os controles, usados para dirigir e regular o veículo. O projeto e a integração desses componentes como um todo são críticos para a dirigibilidade do veículo e determinam o grau de eficiência com a qual os passageiros chegam a seu destino.

Do mesmo modo, um sistema de BPM é uma integração de componentes individuais, cada um dos quais vital por si só, mas quase inúteis se não forem integrados a outros componentes.

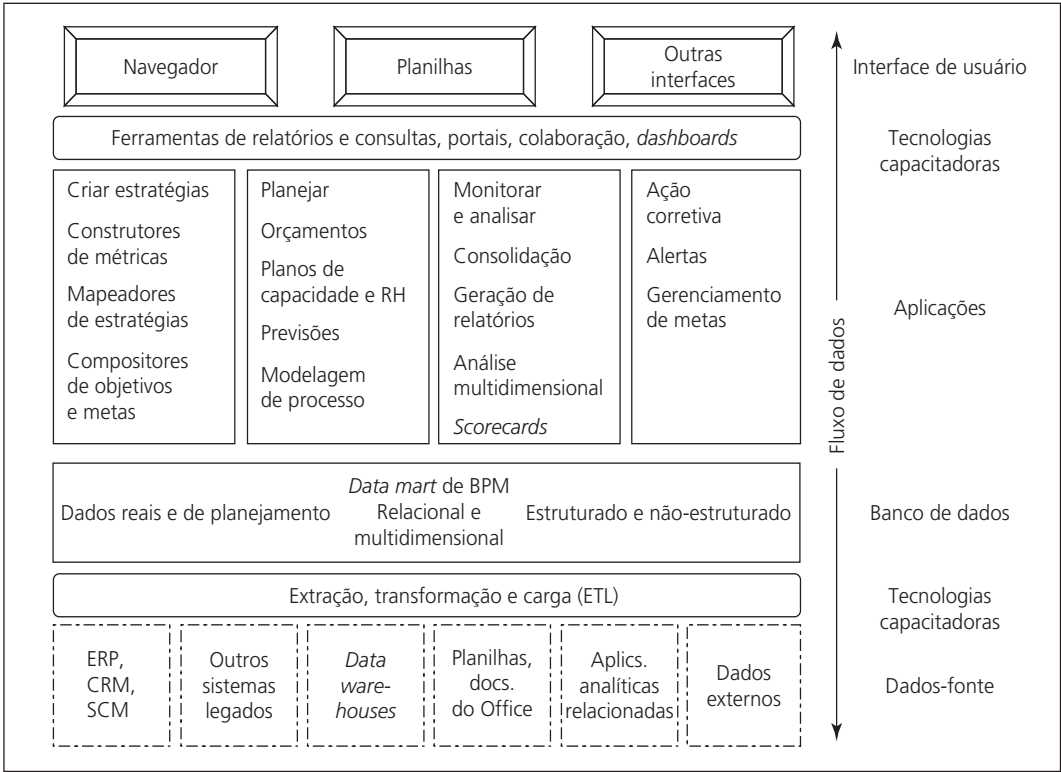


Figura 5.6 Arquitetura de BPM.

Fonte: BPM Standards Group, *Business Performance Management: Industry Framework Document*, September 2005. bpmstandardgroup.org/documents/BPMIndustryFrameworkV5.0090905.pdf (acessado em março de 2006).

Um sistema de BPM necessita três componentes a fim de contribuir para a implementação bem-sucedida da estratégia:

- **Camada de banco de dados.** Esta camada é o chassis de um sistema de BPM. Ela contém uma definição da organização, que abrange seu passado, presente e futuro, e é expressa em termos de estruturas e regras de negócios. Ela também contém informações, ou vínculos a informações, na forma de planos e resultados. É sobre este modelo de dados que os outros componentes de um sistema de BPM operam.
- **Camada de aplicações.** Esta camada é semelhante ao motor de um carro. Ela alimenta os processos de BPM transformando a interação do usuário e os dados-fonte em orçamentos, planos, previsões, relatórios e análises.
- **Camada de cliente ou interface de usuário.** Esta camada é utilizada para dirigir e regular a implementação da estratégia por meio da comunicação, colaboração e orientação da interação do usuário. Ela também monitora o sucesso, a direção e o progresso das iniciativas estratégicas e propicia alertas sobre problemas e oportunidades possíveis e reais.

O ingrediente final são os dados-fonte, que abastecem o sistema de BPM. Em um sistema de BPM, os dados-fonte podem vir de inúmeros sistemas e aplicações internos, como sistemas de planejamento de recursos empresariais (ERP) e *data warehouses*, bem como de fontes externas, como empresas de pesquisa de mercado.

A arquitetura de um sistema de BPM pode afetar radicalmente o que o sistema é capaz de fazer, sua escalabilidade conforme mais usuários ficam *online*, e a quantidade de esforço necessário para a manutenção. Os componentes de uma solução podem ser os melhores do mundo; no entanto, a menos que realmente se encaixem e se complementem bem, o sistema resultante será muito menos que a soma de suas partes.

Camada de banco de dados

Na arquitetura de BPM, a camada de banco de dados proporciona os metadados e os dados sobre os quais se apóiam as aplicações de BPM. Os metadados incluem definições de campo, estruturas hierárquicas, definições de medidas, atribuições de conta, métodos de conversão de moedas, entre outros. Muitos dos dados na camada de banco de dados são estruturados ou quantitativos por natureza e são armazenados em tabelas relacionais ou cubos multidimensionais. Porém, esta camada também armazena outros artefatos de dados que são elementos importantes dos vários processos de BPM. Entre eles, estão incluídos documentos de estratégias, planos táticos, suposições, comentários e alvos de cima para baixo.

A camada de banco de dados em um sistema de BPM pode ser projetada de várias formas. Alguns dos projetos atualmente em uso incluem:

- *Data stores transacionais.* Nesse projeto, a camada de banco de dados não mantém os dados; em vez disso, as aplicações acessam diretamente as fontes de dados transacionais subjacentes, como o razão geral. Apesar de esse projeto fornecer informações muito atualizadas, é provável que resulte em um desempenho ruim, pois as aplicações de BPM terão de acessar e transformar os dados das várias fontes de dados continuamente. Os dados nessas fontes provavelmente serão armazenados em bancos de dados relacionais otimizados para o processamento de transações, não para processamento analítico ou de consultas. Além do mais, não há nenhuma garantia de que os metadados e dados dentro das várias fontes sejam consistentes. Por exemplo, não é garantido que “lucro das vendas” signifique a mesma coisa em todas as várias fontes, ou de que os dados de lucro das vendas sejam os mesmos em todas as fontes.
- *Data marts de aplicações.* Nesse projeto, há *data marts* separados para as várias aplicações de BPM. Os *data marts* são alimentados com dados-fonte provenientes dos sistemas operacionais ou transacionais subjacentes. Esse processo normalmente é realizado com uma ferramenta de extração, transformação e carga (ETL), que é projetada para extrair, transformar e carregar dados de um sistema a outro. Esse projeto tem implementação relativamente rápida e normalmente apresenta um bom desempenho. Diversos fornecedores de BI que oferecem algumas das bases dos sistemas de BPM usam *data marts* desse tipo. Uma desvantagem desse projeto é que administrar e integrar podem tornar-se tarefas difíceis, pois o número de aplicações e *data marts* cresce. Como ocorre com os *data stores* transacionais, também é difícil oferecer uma versão única da verdade em todas as aplicações (p. ex., medidas parecidas podem ter definições e valores diferentes entre *data marts*).
- *Data warehouse centralizado.* Nesse projeto, um banco de dados centralizado armazena a grande maioria dos dados para todas as aplicações de BPM. O projeto tem preferência entre os fornecedores de ERP (como SAP e Oracle) e é usado por alguns fornecedores de BPM. Embora esse projeto garanta uma versão única da verdade, sua implementação (principalmente se iniciada a partir do zero) e manutenção podem ser dispendiosas. Novamente, o *data warehouse* centralizado é alimentado por várias fontes de dados transacionais e operacionais.

A maioria dos sistemas de BPM usa *data marts* ou um *data warehouse*. Os dados em um *data mart* ou *warehouse* são normalmente armazenados em um banco de dados multidimensional ou de processamento analítico *online* (OLAP). O *Multidimensional* descreve a forma como informa-

ções numéricas podem ser categorizadas, visualizadas e analisadas. Um banco de dados multidimensional consiste em uma série de dimensões, como medidas ou contas, tempo, local, produto e versão. Os membros de uma dimensão são organizados hierarquicamente. Em uma dimensão de tempo podem ser organizados como anos, trimestres, meses, semanas e dias. O valor de uma medida dentro do banco de dados é descrito pelos valores dos membros de sua dimensão (p. ex., o lucro real de vendas na região leste no primeiro trimestre).

Em um banco de dados multidimensional, os dados são armazenados em cubos que combinam as várias dimensões dos negócios. Comparado a um banco de dados relacional, um banco de dados multidimensional tem como vantagem seu ótimo desempenho em análises complexas de dados, além de ser relativamente fácil de configurar e manter. Um banco multidimensional também permite que os usuários finais façam visões parciais e cubos dos dados (ou seja, vejam as informações sobre uma determinada combinação de membros de dimensão), realizem *drill up* ou *drill down* na hierarquia (por exemplo, vejam os lucros das vendas por ano, depois por trimestre) e pivotem os dados (por exemplo, vejam os lucros das vendas por produto, depois por local). As desvantagens dos bancos de dados multidimensionais são que se baseiam em números e precisam de tecnologias adicionais para lidar com informações em forma de texto e data, o que é essencial para as soluções de BPM. Eles também não têm padrões, o que significa que muitas aplicações que apresentam bancos de dados multidimensionais são patenteadas. Na melhor das hipóteses, a organização tem de aprender uma nova tecnologia a fim de manter ou ampliar a aplicação. Na pior das hipóteses, isso quer dizer que a organização estará para sempre à mercê do fornecedor do banco de dados para receber atualizações e novas funcionalidades, a fim de permitir que a organização mantenha uma vantagem competitiva com sua estrutura de TI.

A maioria dos dados nos *data marts* ou *warehouses* de BPM não tem origem nos próprios. Em vez disso, os dados são provenientes de uma grande variedade de fontes. Conforme indicado pela Figura 5.7 dos *Insights* de tecnologia 5.5, as fontes podem incluir sistemas operacionais e de cadeia de fornecimento em grande escala (como aplicações de ERP, SCM ou CPM), outras aplicações legadas (como sistemas contábeis), outros *data marts* ou *data warehouses*, planilhas e outras aplicações do Office, aplicações analíticas relacionadas (como o sistema de previsões que alimenta o sistema orçamentário) e dados externos (como dados de pesquisa de mercado). No BPM, as ferramentas de ETL normalmente são utilizadas para mover dados de e para *data marts* e *warehouses* de BPM. O Websphere Data Stage da IBM e o Biztalk da Microsoft são exemplos de ferramentas de ETL.

Aplicações de BPM

No BPM, é necessária uma grande variedade de aplicações para abranger os processos de ciclo fechado, que vão de planejamento estratégico a planejamento operacional e orçamentos a monitoramento a ajustes e ação. Apesar da amplitude dos processos, o grupo Gartner de análises do setor afirma que as seguintes aplicações podem lidar com a maioria dos processos (consulte RAYNER et al.: 2005a):

1. *Orçamentos, planejamento e previsões.* Essas aplicações suportam o desenvolvimento de todos os aspectos de orçamentos, planos e previsões. Elas englobam orçamentos de curto prazo focados em finanças, planos de longo prazo e planos estratégicos de alto nível. As aplicações devem proporcionar recursos de fluxo de trabalho para gerenciar a criação, envio e aprovação de orçamentos/planos, e oferecem facilidade para criar previsões e cenários de forma dinâmica. Elas também devem suportar o desenvolvimento de um modelo de planejamento para toda a empresa que vincule os planos operacionais aos orçamentos financeiros. Além disso, devem ser capazes de compartilhar dados com aplicações de domínios específicos, como planejamento da cadeia de fornecimento.
2. *Modelagem e otimização da lucratividade.* Essas aplicações incluem aplicações de custeio baseado em atividades (ABC) que determinam e alocam os custos a um nível altamente gra-

- nular. Elas incluem também aplicações de gestão baseadas em atividades que propiciam recursos para permitir aos usuários modelar o impacto de diferentes estratégias de alocação de custos e recursos sobre a lucratividade. Algumas aplicações foram além do tradicional foco em ABC e permitiram que o lucro fosse alocado, além de modelos de custos de empacotamento, agrupamento, precificação e estratégias de canal.
3. *Aplicações de scorecards.* *Scorecards* são um recurso genérico de BI que também podem vincular indicadores de desempenho a um mapa estratégico com uma relação hierárquica de causa e efeito entre os KPIs. Os *scorecards* são frequentemente usados em conjunto com uma determinada metodologia, como BSC ou Six Sigma.
 4. *Consolidação financeira.* Esse tipo de aplicação permite que as organizações conciliem, consolidem, resumam e agreguem dados financeiros com base em diferentes padrões contábeis e regulamentos federais. Essas aplicações são uma parte fundamental do CPM, pois criam a visão auditada, no âmbito da empresa, das informações que devem ser compartilhadas com outras aplicações de CPM para analisar o desvio das metas.
 5. *Relatórios estatutários e financeiros.* As aplicações de BPM exigem ferramentas especializadas de geração de relatórios que possam formatar os dados de saída como relatórios financeiros estruturados. Elas podem também necessitar suportar regras de apresentação de princípios contábeis geralmente aceitos (GAAP), como as GAAP norte-americanas ou padrões de relatórios financeiros internacionais. Elas incluem também técnicas de visualização projetadas especificamente para oferecer suporte à análise do desvio de orçamentos ou metas, como árvores hiperbólicas.

Interface de usuário do BPM

A interface de usuário é a ponte entre as aplicações de BPM e o usuário final. A interface particular fornecida depende da aplicação específica que está sendo acessada, bem como do papel do usuário, seus objetivos e sua experiência. Por exemplo, um gerente, que é usuário ocasional e está interessado em revisar os resultados, provavelmente vai precisar e usar uma interface diferente daquela do gerente operacional, que é usuário frequente e tem interesse em investigar vários cenários táticos ou orçamentários.

Nos últimos anos, o navegador da Web tornou-se a ferramenta principal de acesso às informações em um sistema de BPM. Sistemas baseados na Web são extremamente eficientes quando se trata de uma grande implantação por toda a empresa. Não há necessidade de carregar ou manter um *software* em uma máquina. O usuário final precisa simplesmente de um endereço da Web e um navegador para acessar a aplicação. A Web permite a realização contínua de mudanças no sistema de BPM, sem exigir um processo penoso de implementação. Todas as alterações feitas a um sistema baseado na Web são instantaneamente comunicadas a todos os usuários, sem a necessidade de enviar arquivos, templates ou qualquer outra coisa. A Web também oferece acesso a uma variedade de dados: texto, números formatados, imagens, vídeos, sons, documentos, entre outros. Isso quer dizer que ela pode ser usada com uma grande variedade de aplicações.

Para alguns usuários e algumas aplicações, um navegador da Web não é o suficiente. Isso se dá principalmente com aplicações de planejamento e previsões, nas quais é necessária uma interface de usuário elaborada para dar suporte às necessidades analíticas e computacionais do usuário. Nesses casos, as planilhas se tornaram uma alternativa popular. Aqui, o programa de planilhas, como o Excel, serve como uma janela para a aplicação e como o banco de dados de BPM de retaguarda *back-end*. Os dados da aplicação aparecem diretamente nas células da planilha. O usuário final pode tirar proveito dos recursos de cálculos, relatórios e gráficos das planilhas. Em muitos casos, isso também permite que o usuário final trabalhe *offline* após desconectar-se da aplicação e do banco de dados de suporte, e envie alterações à aplicação após conectar-se novamente.

Algo que todas as interfaces de BPM deveriam oferecer é orientação ao usuário final. Posto que há tantos dados e alguns dos processos são tão complexos, é muito fácil que um usuário final se afunde em uma sobrecarga de dados e se confunda tentando navegar por um sistema. Essa é uma das razões pelas quais os recursos de BI, como exceções identificadas por cores, alertas de exceção, listas ordenadas, etc., são tão valiosos. Sem esses tipos de componentes de interface de usuário, um usuário final poderia passar horas peneirando os dados, tentando identificar desvios e tendências significativos. Esses tipos de recursos são fornecidos por *scorecards* e *dashboards* de BPM (veja a Seção 5.10).

Suítes comerciais de BPM

Segundo estimativas do Gartner, o mercado de *software* comercial de suítes de BPM era de aproximadamente US\$ 500 milhões em 2003, e a previsão é de que esteja em torno de US\$ 1 bilhão em 2009 (Rayner et al.: 2005b). Em contraste, a firma de pesquisas International Data Corporation (IDC) estimou que o mercado de aplicações de BPM era de US\$ 1,2 bilhões em 2003 e esperava que crescesse atingindo US\$ 2,1 bilhões em 2008 (WHITING: 2005). O mercado é composto pelas empresas de *software* que oferecem suítes com ao menos três das principais aplicações fundamentais de BPM (ou seja, orçamentos, planejamento e previsões; modelagem e otimização da lucratividade; *scorecarding*; consolidação financeira; relatórios estatutários e financeiros). Devido à taxa de crescimento anual composta (CAGR) do mercado ser de cerca de 10%, o número de novos participantes no mercado cresceu no último ano. Ao mesmo tempo, houve diversas aquisições feitas por companhias de *software* de BPM que buscavam expandir suas ofertas de aplicações de BPM.

Assim como faz com vários mercados de *software* que acompanha, o Gartner estabeleceu um *quadrante mágico* para os fornecedores de suítes de CPM (leia Rayner et al.: 2005b). O quadrante posiciona as empresas com base em sua capacidade de execução (como empresa) e na plenitude de suas visões. A combinação das duas dimensões resulta em quatro categorias de empresas. Os fornecedores de *software* dentro de cada uma das categorias são apresentados na Tabela 5.4.

Tabela 5.4 Quadrante mágico das suítes de BPM

Visão		
Execução	Limitação	Força
Força	Desafiadores: SAP	Líderes: Hyperion Solutions, Cognos
Limitação	Nicho: Oracle/PeopleSoft EPM, Lawson Software, Applix, Systems Union, Coda, Clarity Systems, CorVu	Visionários: Oracle CPM, Business Objects, SAS Institute, OutlookSoft, Cartesis, Longview Solutions, ALG Software

Fonte: N. Rayner et al., "Magic Quadrant for CPM Suites, 2005". Gartner RAS Core Research NoteG00131208, October 2005.

Com base nos *rankings* do Gartner, os líderes absolutos no mercado de suítes de BPM são a Hyperion Solutions e a Cognos. A discussão dos *Insights* de tecnologia 5.5 oferece uma visão detalhada sobre uma das suítes de BPM líderes de mercado: Hyperion System 9.

Questões de revisão da Seção 5.9

- 1. O que é uma arquitetura de sistema?
- 2. Quais são os três elementos-chave em uma arquitetura de sistema de BPM?

INSIGHTS DE TECNOLOGIA 5.5

A estrutura e os benefícios de uma suíte comercial de BPM

A Hyperion usa o termo *BPM* para descrever suas ofertas de gerenciamento do desempenho. Sua suíte de BPM chama-se Hyperion System 9. O Hyperion System 9 combina a plataforma de BI da Hyperion com sua suíte de aplicação financeira. Ele foi projetado e construído com o uso de uma arquitetura orientada a serviços (SOA) moderna e em *n* camadas. Como mostra a Figura 5.7, a arquitetura SOA é semelhante à estrutura arquitetônica oferecida pelo BPM Standards Group.

ARQUITETURA DO SISTEMA

Conceitualmente, o Hyperion System 9 consiste em quatro camadas:

1. *Cliente*. Aos usuários finais, o System 9 oferece um ambiente rico e baseado em servidor via HTML dinâmico (DHTML). Ele proporciona aos usuários de negócios um ponto único de acesso a todo o conteúdo de BPM, que pode ser personalizado de acordo com suas necessidades individuais. Oferece também o Smart View, um suplemento do Microsoft Office que permite que os usuários de negócios trabalhem diretamente no ambiente do Office (ou seja, Excel, Word, PowerPoint, Outlook) enquanto interagem ao vivo com seu conteúdo de BPM da Hyperion.
2. *Web*. A camada da web do Hyperion System 9 usa os servidores de aplicação Web Java 2 Enterprise Edition (J2EE), como o BEA WebLogic e o IBM WebSphere. Ela é composta por um conjunto de aplicações web construídas como Java *servlets* e páginas JSP. Essas aplicações Web orientadas por objetivos incorporam lógica específica ao tipo de funcionalidade de BPM a ser oferecido, além de propiciarem recursos altamente escalonáveis.
3. *Serviços*. Os serviços que fornecem funcionalidades de BPM especializadas são construídos segundo os objetivos, a fim de oferecer conteúdo para os recursos analíticos, de relatórios, de aplicações e de acesso a dados do Hyperion

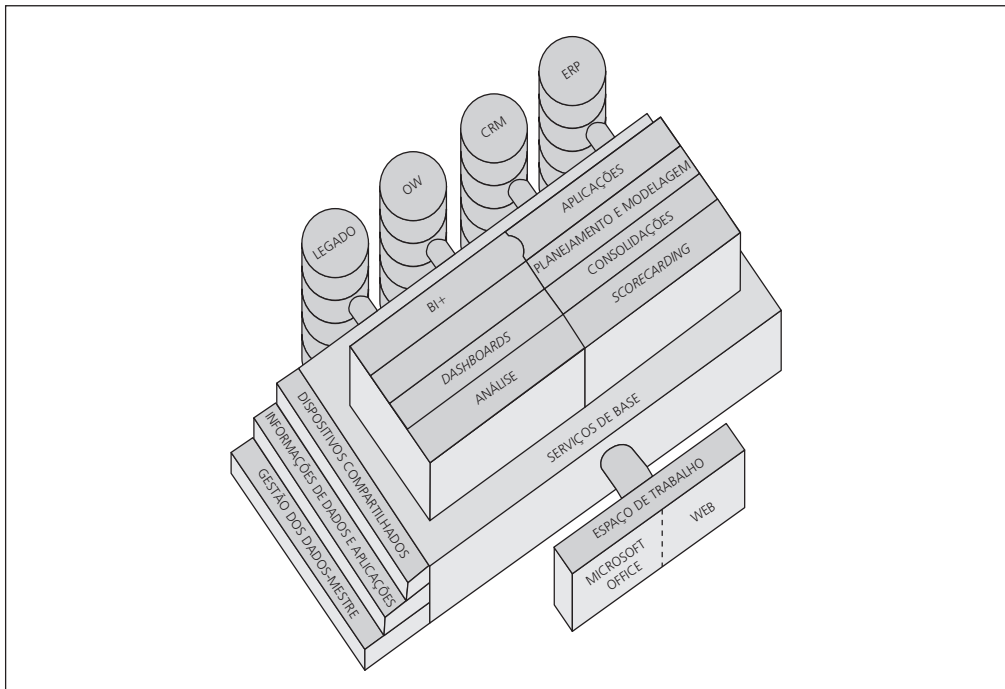


Figura 5.7 Arquitetura de BPM da Hyperion Solutions.

Fonte: hyperion.com/products/.

(Continua)

INSIGHTS DE TECNOLOGIA 5.5 (continuação)

System 9. Essas funcionalidades incluem o serviço Financial Reporting, para a produção de dados de relatórios orientados a finanças; o serviço Production Reporting, para proporcionar conteúdo em tela de relatórios operacionais; o serviço Data Access, usado pelo módulo Interactive Reporting para acessar bancos de dados relacionais; o serviço Interactive Reporting, para gerenciar *dashboards*, consultas, pivôs, gráficos e relatórios de gestão; e o sofisticado serviço Web Analysis, para navegar por dados em OLAP multidimensional.

4. **Dados.** O Hyperion System 9 unifica relatórios, consultas e análises, e dissemina aplicações financeiras por uma grande variedade de fontes de dados. A suíte suporta acesso a inúmeras fontes relacionais, legadas e multidimensionais. As aplicações do System 9 se utilizam da tecnologia adequada de banco de dados para oferecer a funcionalidade de negócios necessária exigida pela aplicação. Por exemplo, o Hyperion Planning explora uma estrutura de dados multidimensional para armazenar seus dados, a fim de otimizar os recursos de atualização e regras de negócios exigidos pelos usuários de planejamento. Enquanto isso, os metadados e as anotações são armazenados em um banco de dados relacional.

APLICAÇÕES DO HYPERION SYSTEM 9

O Hyperion System 9 suporta uma ampla variedade de aplicações financeiras e de relatórios de BI. Entre as aplicações financeiras de BPM oferecidas pela suíte estão a modelagem financeira (Strategic Finance); planejamento, orçamentos e previsões financeiras baseados na Web (Planning); planejamento de número de funcionários, salários e compensações (Workforce Planning); e *software* de sistemas financeiros com base na Web para proporcionar relatórios e análises de coleta global (Financial Management). Além dessa gama de funcionalidades financeiras, o Hyperion System 9 também proporciona uma variedade de aplicações de relatórios de BI, incluindo Interactive Reporting, para consulta e análise direcionadas ao usuário; Financial Reporting, para gerar relatórios financeiros com qualidade de livros contábeis, que cumprem com normativas e exigências externas; Production Reporting, para gerar relatórios formatados, de alto volume e qualidade de apresentação;

Compliance Management *Dashboard*, para combinar informações de controle interno com dados financeiros a fim de propor uma avaliação das possíveis exposições a riscos; e Enterprise Metrics, que é um mecanismo e biblioteca centralizado de métricas que reduz o tempo e as despesas com criação, medição e geração de relatórios de KPIs.

CUSTO TOTAL DE PROPRIEDADE (TCO)

À medida que as empresas lutam para fazer uma melhor gestão de suas despesas, a TI procura diminuir o TCO de todos os sistemas gerenciados. O Hyperion System 9 cumpre com este objetivo das seguintes formas:

- O Hyperion System 9 é uma abordagem unificada ao BPM, incorporando amplitude de funções às ferramentas de BI e às aplicações financeiras: tudo em um único pacote de produtos. Com um sistema a ser licenciado e um fornecedor com quem estabelecer parceria, a empresa pode suprir todas as suas necessidades de BPM e chegar a um TCO inferior.
- O Hyperion System 9 é construído com o uso de princípios de SOA e tecnologias modernas, como DHTML, CSS, XML, JavaScript e J2EE. Ele adota o valioso paradigma de aplicações da Internet utilizando-se de técnicas de AJAX, oferece uma incomparável experiência de usuário e é adequado aos padrões exigentes instituídos pelas atuais organizações de TI.
- O Hyperion System 9 pode começar sendo pequeno e crescer com as necessidades analíticas da firma, ou ainda ser escalonado rapidamente por toda a empresa. A filosofia "comece de qualquer lugar" da Hyperion permite que as empresas priorizem suas necessidades, instalem apenas os serviços necessários e depois implantem as soluções conforme for preciso. Além disso, sua arquitetura multicamadas permite implementações muito avançadas.
- O Hyperion System 9 é fácil de usar, gerenciar e manter. Devido à sua aparência e comportamento comum, os usuários necessitam de treinamento somente uma vez e depois podem se atualizar rapidamente conforme novas aplicações são disponibilizadas. Pontos centrais de administração aliviam o peso sobre os ombros da TI.

3. O que um banco de dados de BPM oferece?
4. Quais são alguns projetos alternativos de banco de dados de BPM?
5. Quais são as principais categorias de aplicações de BPM?
6. Quais são alguns dos tipos de interfaces de usuário de BPM?
7. Quais são as categorias básicas do quadrante mágico do Gartner? Quem são alguns dos fornecedores nas diversas categorias?
8. Quais são algumas das aplicações de BPM que a Hyperion Solutions suporta?

5.10

DASHBOARDS DE DESEMPENHO

Scorecards e *dashboards* são componentes comuns, se não de todos, da maior parte dos sistemas de gerenciamento do desempenho, sistemas de medição do desempenho e suítes de BPM. Tanto *dashboards* quanto *scorecards* proporcionam exibições visuais de informações importantes, que são consolidadas e organizadas em uma tela única para serem absorvidas em uma passada de olhos e exploradas facilmente. Um *dashboard* típico é exibido na Figura 5.8. Este *dashboard* em especial exibe vários dados de KPI e *pipeline* para uma empresa de *software*, que produz componentes especializados de gráficos e apresentações visuais para desenvolvedores de *software* e os vende diretamente pela Web. A partir do *dashboard*, é fácil ver, por exemplo, que todos os KPIs estão bons (estão todos no verde, embora apareçam em cinza neste

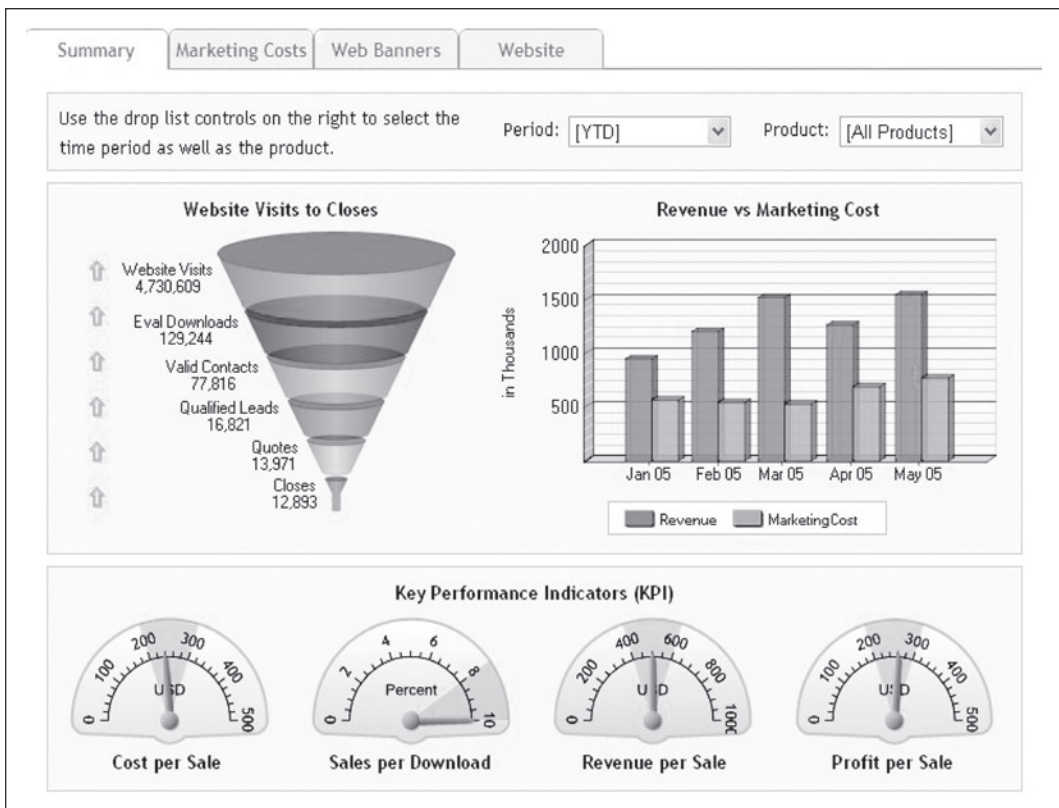


Figura 5.8 Exemplo de *dashboard* de desempenho.

Fonte: Dundas Software, demos1.dundas.com/DundasGauge/MarketingDashboard/WebSite.aspx.

livro); que, para todos os estágios do *pipeline*, os números tendem a subir (todos são setas verdes apontando para cima, apesar de estarem na cor cinza neste livro); e que o crescimento das vendas está superando o aumento nos custos de *marketing*. Este *dashboard* em especial permite que os usuários finais vejam se há diferenças por período de tempo ou por produto (os menus suspensos na parte superior direita) e analisem melhor os custos de *marketing* e o *pipeline* da Web (selecionando as diferentes guias na parte superior).

Dashboards versus scorecards

Ainda que *dashboards* e *scorecards* tenham muito em comum, existem diferenças entre os dois. Algumas diferenças estão resumidas na Tabela 5.5. A principal delas é que os ***dashboards de desempenho*** são exibições visuais usadas para monitorar o desempenho operacional, enquanto os ***scorecards de desempenho*** são exibições visuais usadas para mapear o progresso com relação a metas e alvos estratégicos e táticos.

Nas revistas especializadas, com frequência os dois termos são usados indiscriminadamente, apesar de muitos fornecedores de CPM e BI oferecerem soluções e aplicações separadas de *dashboards* e *scorecards*. A Tabela 5.6 oferece uma lista com algumas dessas ofertas comerciais.

Na maioria das ofertas comerciais, o termo *scorecard* é praticamente um sinônimo do conceito de um BSC. Na verdade, a maioria dos produtos de *scorecard* de BPM têm o certificado do BSC Collaborative, o que significa que todos têm características de exibição relativamente similares. Por exemplo, todos oferecem a aparência de um mapa estratégico.

Eckerson, uma das principais autoridades em BI em geral e em *dashboards* de desempenho em particular, sugeriu que um termo se encaixa em todas as situações: *dashboards* de desempenho. Em seu jargão, um *dashboard* de desempenho é “uma aplicação em várias camadas construída sobre uma infra-estrutura de *business intelligence* e integração de dados, que permite às organiza-

Tabela 5.5 Diferenças entre *dashboards* e *scorecards* de desempenho

<i>Característica</i>	<i>Dashboard</i>	<i>Scorecard</i>
Propósito	Mede o desempenho	Mapeia o progresso
Usuários	Supervisores, especialistas	Executivos, gerentes, equipe
Atualizações	Transmissões “na hora certa”	Instantâneos periódicos
Dados	Eventos	Resumos
Exibição	Gráficos visuais, dados brutos	Gráficos visuais, comentários em texto

Fonte: W. Eckerson, *Performance Dashboards*, Wiley, Hoboken, NJ, 2006.

Tabela 5.6 Produtos de *software* comercial de *dashboards* e *scorecards*

<i>Fornecedor</i>	<i>Produto de dashboard</i>	<i>Produto de scorecard</i>
Cognos	Cognos 8 BI Dashboards	Cognos 8 BI Scorecarding
Hyperion	Hyperion Compliance Management Dashboard, Hyperion System 9 BI+ Enterprise Metrics	Hyperion System 9 Performance Scorecard
SAS	SAS Strategic Performance Management	SAS Strategic Performance Management
Oracle	Oracle PeopleSoft Operational Dashboards	Oracle Balance Scorecard
Information Builders	WebFocus Business Intelligence Dashboard	WebFocus Performance Management Framework

ções medir, monitorar e gerenciar o desempenho dos negócios com mais eficiência” (ECKERSON: 2006). Em seu esquema, diferenciam-se três tipos de *dashboards* de desempenho:

1. *Dashboards operacionais*. Usados por funcionários da linha de frente e por supervisores para monitorar os principais dados operacionais que são ligeiramente resumidos e atualizados com frequência durante o dia.
2. *Dashboards táticos*. Usados por gerentes e analistas para acompanhar diária ou semanalmente dados detalhados e resumidos, gerados a partir de processos e projetos departamentais. O sistema no pronto-socorro discutido no Caso de aplicação 5.6 exemplifica este tipo de *dashboard*.
3. *Dashboards estratégicos*. Usados por executivos, gerentes e equipe para monitorar mensal ou trimestralmente dados detalhados e resumidos, pertencentes à execução de objetivos estratégicos.

CASO DE APLICAÇÃO 5.6

Dashboards para médicos

Uma das forças motrizes por trás da aceitação de *dashboards* de desempenho é o desejo de colocar funcionalidades de BI nas mãos de usuários que não sejam analistas de negócios ou usuários avançados. Um caso pertinente é o do Emergency Medicine and Analysis Report System (EMARS), desenvolvido por médicos e usado pelo Emergency Medical Associates (EMA), um grupo médico sem fins lucrativos que administra prontos-socorros de 16 hospitais em Nova York e Nova Jersey. Como afirmou Jonathan Rothman, diretor de gestão de dados do EMA, “o *dashboard* realmente é o melhor [método] para o departamento de emergências (DE). Os velocímetros são ótimos para uma análise rápida dos dados, pois, uma vez que você entende a métrica e compreende quais são os alvos e limites dos usuários finais, o resto é basicamente olhar as métricas e procurar por algo em vermelho”.

INCENTIVO

O EMA procurava uma forma mais intuitiva dos médicos acessarem as informações armazenadas em seu *data warehouse* Oracle. Antes da introdução do sistema de *dashboards*, a única maneira dos médicos obterem acesso aos dados no *warehouse* era por um sistema legado de acompanhamento de pacientes e registros médicos eletrônicos. O sistema legado, chamado Emergency Department Information Manager (EDIM), havia sido construído no Visual FoxPro e oferecia diversos modelos clínicos e relatórios predefinidos. Apesar dessas capacidades incorporadas, o sistema era difícil de usar e de ampliar.

Para superar as limitações do sistema EDIM, o EMA recorreu ao Application Foundation *Dashboard Manager* da Business Objects, um *kit* de ferramentas de análise que permite aos desenvolvedores de aplicações criar *dashboards* de análise detalhados com um mínimo de esforço. O Application Foundation possibilita que os desenvolvedores apresentem métricas de alto nível em formato de *dashboard* e as conectem aos detalhes operacionais subjacentes. Dessa maneira, os usuários finais podem fazer *drill down* até os detalhes de qualquer métrica de alto nível que exiba padrões de interesse.

MÉTRICAS DE INTERESSE

Que tipos de métricas são do interesse dos médicos e administradores do DE? Não as informações médicas em si. Em vez disso, os médicos e administradores do DE querem dados que os ajudem a monitorar a saúde do próprio departamento. Total de lucros, total de gastos, número de consultas com pacientes e porcentagem geral de internações no DE são algumas das métricas disponíveis nos *dashboards* do EMARS. Basicamente, os *dashboards* dão aos médicos e outros funcionários uma visão estatística do “bem-estar” do pronto-socorro, em termos financeiros e de serviço.

Há dois tipos de *dashboards* no EMARS: os associados a faturamento e arrecadamento e os associados a serviços do departamento de emergências. Os *dashboards* de faturamento e arrecadamento são bastante objetivos e oferecem informações sobre lucros e despesas. Esses são os tipos de métricas encontradas nos *dashboards* de muitas organizações. Por outro lado, as métricas de serviço são muito específicas às operações

de pronto-socorro. Tomemos como exemplo uma métrica de desempenho chamada “saiu sem ser examinado” (LWOB). No jargão hospitalar, isso quer dizer que o paciente foi ao pronto-socorro, esperou e ao final foi embora do lugar, possivelmente para procurar ajuda em outro local. Informações como essa ajudam o EMA a ter mais sucesso. Se um pronto-socorro em especial tem um LWOB que supera um limite predeterminado, isso significa que o serviço está lento e que são necessárias mudanças. O mesmo é válido para outras métricas de serviço. Algumas das outras métricas que o EMARS acompanha incluem total de consultas pediátricas, total de internações de pacientes e percentual de pacientes que saíram contra parecer médico.

Além dessas métricas de serviço de alto nível, o EMARS também proporciona informações mais granulares. Um exemplo é o tempo que leva entre um paciente fazer seu cadastro e ser propriamente atendido pelo médico. Outros exemplos são o tempo que os pacientes internados levam para receber alta e o tempo que os pacientes que receberam alta levam para voltar.

Muitas das métricas de serviço têm referenciais nacionais. Isso permitiu que o EMA predefinissem li-

mites para muitos deles. Quando um limite é superado, envia-se um alerta por *e-mail* aos médicos e administradores interessados. O alerta aponta não só a exceção, mas também os tipos de análise que podem ajudar o destinatário a compreender por que a exceção ocorreu.

Em uma perspectiva de TI, uma das desvantagens dos *dashboards* é que o *data warehouse* Oracle por trás de tudo tem de ser alimentado a partir de diversas fontes externas por uma ferramenta de ETL. Algumas das fontes principais são, na verdade, repositórios do Crystal Reports. (Crystal Reports é de propriedade da Business Objects.) Futuramente, o departamento de TI poderia pular esta etapa e simplesmente publicar o Crystal Reports diretamente no *dashboard*. A Business Objects indicou que pretende oferecer esse recurso no futuro.

Fontes: compilado de D. Bartholomew, “Gauging Success with *Dashboards*, What You See Is Only Part of What You Get”, *CFO IT*, August 2005, cfo.com/article.cfm/4077408/c_3242196 (acessado em março de 2006); and S. Swayer, “*Dashboards* for Doctors”, *BI Intelligence Journal*, Fall 2004.

Projeto do *dashboard*

Os *dashboards* englobam muitas informações em uma única tela. “O desafio fundamental do projeto do *dashboard* é exibir todas as informações claramente, em uma única tela e sem distrações, de uma forma que possa ser assimilada rapidamente” (FEW: 2005). O Arquivo *online* W5.5 detalha parte desse desafio. Em sua maioria, os *dashboards* exibem medidas quantitativas sobre o que está acontecendo. Para acelerar a assimilação dos números, estes precisam ser colocados em um contexto. Isso pode ser feito comparando os números relevantes a outros números de referência ou de alvo, indicando se os números estão bons ou ruins, mostrando se uma tendência é positiva ou negativa, e usando mecanismos ou componentes especializados de visualização para definir o contexto comparativo e evolutivo.

Uma forma de colocar os números em um contexto é por meio da comparação. Sozinho, um número não significa muito. Se alguém diz que o lucro das vendas de uma empresa foi de 20 milhões no último trimestre, é difícil determinar o que isso quer dizer em termos de desempenho da empresa. Porém, se alguém diz que a empresa fez uma previsão de 25 milhões para o trimestre, ou que no ano passado faturou 30 milhões no mesmo trimestre, isso coloca os números em uma situação diferente. Dadas essas comparações, provavelmente se supõe que o desempenho não é o esperado pela empresa. Algumas das comparações normalmente feitas em um sistema de BPM incluem comparações com valores passados, valores previstos, valores-alvo, valores médios ou referenciais, múltiplas ocorrências da mesma medida e valores de outras medidas (p. ex., lucro vs. custo). Na Figura 5.8, os vários KPIs são colocados em contexto por meio de comparações com custos de *marketing*, e os números dos vários estágios do *pipeline* de vendas são colocados em contexto através de comparações entre os estágios.

Mesmo com medidas comparativas, é importante apontar especificamente se um determinado número está bom ou ruim e se sua tendência vai na direção certa. Sem esse tipo de designação

avaliativa, determinar o status de um número ou resultado pode ser demorado. Geralmente, objetos visuais especializados (como semáforos) ou atributos visuais (como códigos de cores) são usados para definir o contexto avaliativo. Ainda sobre o *dashboard* da Figura 5.8, usam-se códigos de cores com os indicadores para designar se os KPIs estão bons ou ruins; setas verdes apontando para cima são usadas com os vários estágios do *pipeline* de vendas para indicar se os resultados tendem a subir ou a descer, e se subir ou descer é bom ou ruim. Embora não sejam usados neste exemplo em especial, outras cores, como vermelho e laranja, poderiam ser usadas para representar outros estados avaliativos nos vários indicadores.

O que procurar em um *dashboard*

Embora haja diferenças entre *dashboards* de desempenho e *scorecards*-padrão de desempenho, esses dois sistemas compartilham algumas das mesmas características. Primeiro, ambos se encaixam no sistema maior de BPM ou de medição do desempenho. Isso quer dizer que sua arquitetura subjacente é a arquitetura de BI ou de gerenciamento do desempenho do sistema maior. Segundo, todos os *dashboards* e *scorecards* bem planejados possuem as seguintes características:

- Usam componentes visuais (p. ex., gráficos, barras de desempenho, *sparklines*, indicadores, medidores, semáforos) para destacar de forma imediata os dados e exceções que exigem ação.
- São transparentes ao usuário; ou seja, exigem treinamento mínimo e são extremamente fáceis de usar.
- Combinam dados de diversos sistemas e formam uma visão dos negócios única, resumida e unificada.
- Possibilitam a realização de *drill down* (ou navegar através) em fontes de dados ou relatórios, oferecendo mais detalhes sobre o contexto comparativo e avaliativo que está por trás.
- Apresentam uma visão dinâmica e prática com atualizações pontuais de dados, o que permite ao usuário final estar atualizado sobre quaisquer alterações recentes nos negócios.
- Exigem poucos, ou nenhum, códigos customizados para implementar, implantar e manter.

Questões de revisão da Seção 5.10

1. Quais são as principais diferenças entre um *scorecard* e um *dashboard*?
2. O que é um *dashboard* operacional? E um *dashboard* tático? E um *dashboard* estratégico?
3. O sistema EMARS possuía que tipo de *dashboard*: operacional, tático ou estratégico? Por que este tipo era útil para os médicos e administradores do pronto-socorro?
4. Que critérios são importantes ao selecionar os mecanismos de visualização específicos para usar com determinadas métricas em um *dashboard*?
5. Quais são as características de um *dashboard* bem planejado?

5.11

MONITORAMENTO DE ATIVIDADES DE NEGÓCIOS (BAM)

Monitoramento de atividades de negócios (BAM) é um termo inventado pelo Gartner. O termo reflete o interesse deles e de outros nos conceitos estratégicos da empresa de latência zero e do processamento direto (*straight-through processing*) (relatado por McKie, 2003; e Systar, 2002). Nesse contexto, *latência* refere-se à lacuna entre o momento em que os dados são coletados e o momento em que estão disponíveis para tomada de decisão. Uma *empresa de latência zero* é aquela na qual os dados estão imediatamente disponíveis, permitindo a uma empresa ser proativa ao invés de reativa. *Processamento direto* refere-se ao processo no qual etapas ineficientes (como registro manual) são eliminadas.

Sistemas BAM consistem em sistemas em tempo real que alertam os gerentes sobre possíveis oportunidades, problemas iminentes ameaças, e os habilitam a reagir usando modelos e colabo-

ração. A TI pode coletar dados, em tempo real, de uma variedade de fontes internas e externas, analisá-los para detectar padrões inesperados que indicam uma situação em evolução e, então, fornecer os resultados para os responsáveis pela reação. Esse aspecto tecnicamente ágil da *empresa em tempo real* muitas vezes é rotulado por *BAM* (veja Keating, 2003).

Os principais usuários de BAM incluem executivos de linha de negócios, gerentes de departamento, equipes de operações comerciais e CFOs. Normalmente, BAM é usado para supervisionar as atividades de uma instalação específica, como uma fábrica ou uma central de atendimento, ou um processo de negócios específico. O Caso de aplicação 5.7, que descreve o uso de BAM feito pelo governo da cidade de Albuquerque, ilustra a natureza orientada a processo e voltada para um único propósito do BAM.

CASO DE APLICAÇÃO 5.7

A cidade de Albuquerque em tempo real

Em uma pesquisa de 2003 realizada pela Ventana Research, foi solicitado que 1.300 gerentes de negócios e de TI indicassem as três metas mais importantes para monitoramento dos seus negócios. As principais razões dadas foram: administrar ou reduzir custos (29%), melhorar a eficiência da empresa (29%), aumentar o foco nas prioridades de receita (19%), alinhar ações individuais às prioridades (15%) e reagir contra os concorrentes (5%). Como era de se prever, eficiência foi considerada mais importante do que efetividade.

Eficiência é a principal razão para a cidade de Albuquerque ter instituído um sistema BAM. O governo da cidade conta com o NoticeCast da Cognos, uma proeminente fornecedora de *software* de BI e BPM, para acelerar proativamente avisos de *e-mail* de eventos importantes quase em tempo real para funcionários, moradores e fornecedores da cidade. O NoticeCast tem a habilidade de monitorar três categorias básicas de eventos:

- *Eventos de notificação.* O sistema notifica os usuários toda vez que um relatório, cubo OLAP ou uma consulta é atualizado.
- *Eventos de desempenho.* O sistema notifica os usuários toda vez que um limite atual ou projetado é ultrapassado.
- *Eventos operacionais.* O sistema notifica os usuários toda vez que uma transação específica acontece nas operações diárias.

O NoticeCast também possibilita aos usuários finais criar seus próprios eventos, fornece entrega flexível de alertas a dispositivos com e sem fio, e integra-se aos pacotes de BPM e BI da Cognos.

No sistema de Albuquerque, o NoticeCast está em uma *extranet* fora do *firewall* da cidade, e monitora os eventos através de consulta periódica às tabelas

do banco de dados Oracle que são alimentados periodicamente e muitas vezes automaticamente por outro sistema municipal de informações. Os tipos de eventos monitorados pelo sistema incluem (Anthes, 2003):

- A cada manhã, o NoticeCast envia um *e-mail* a cada fornecedor para o qual foi emitido um pagamento eletrônico durante a noite, direcionando o fornecedor a um *website* na *extranet* onde pode obter um relatório de remessa.
- A cada anoitecer, o NoticeCast envia um *e-mail* a cada morador de Albuquerque para o qual uma conta de água foi gerada. O *e-mail* contém todas as informações de faturamento pertinentes e direciona o morador ao *website* onde ele/ela pode pagar sua conta *online*.
- Uma vez por dia, o sistema envia um *e-mail* para determinados funcionários do município, informando-os de todos os pagamentos *online* feitos pela cidade nas últimas 24 horas.
- Toda vez que um candidato registra um relatório de contribuição, o NoticeCast envia um *e-mail* aos funcionários do município responsáveis pelo controle do cumprimento da lei eleitoral.

O sistema BAM ajuda a cidade a monitorar eventos de modo mais rápido e mais criterioso do que antes. Também lhes permite realizar uma série de atividades, como faturamento *online*, que eram difíceis ou muito caras no passado. No futuro, a cidade planeja implantar o sistema dentro do seu *firewall*, na sua *intranet*.

Fontes: compilado de G. Anthes, "Eyes Everywhere: Business Activity Monitoring Offers a Constant Watch on Business Processes", *Computerworld*, November 2003, computerworld.com/printthis/2003/0,4814,86895,00.html (acessado em março de 2006); and cognos.com.

BAM depende de uma ampla variedade de tecnologias trabalhando em harmonia. Os seguintes itens são de especial importância:

- *Tecnologia ETL* reúne dados de múltiplas fontes que podem ser relevantes para uma atividade de aplicação cruzada visualizada pela:
- *Tecnologia de modelagem de processo* define o escopo de atividades relevantes e identifica etapas distintas do processo naquelas atividades que estão sujeitas aos:
- *Mecanismos de regras* definem e aplicam as regras que permitem que eventos significativos dentro dessas atividades sejam reconhecidos e respondidos usando:
- *Servidores de mensagem* agrupam e comunicam ou propagam a resposta para esses eventos em vários formatos a fim de funcionar como um fornecedor de:
- *Caixas de entrada de e-mails, portais, dashboards e web services* usam essas respostas para revelá-las aos tomadores de decisão que podem se guiar por elas ou sistemas que podem processá-las no futuro (talvez para acionar outro evento ou um fluxo de trabalho completo).

Benefícios do BAM

O BAM ajuda não apenas no reconhecimento e na resposta a eventos, mas também na possibilidade de gerentes resolverem ocorrências de eventos rapidamente e analisarem suas influências para tomarem decisões mais informadas e em tempo hábil (veja McKie, 2003). Essencialmente, os dois benefícios mais importantes são acesso em tempo real aos dados em um formato utilizável e acesso às ferramentas para colaboração e modelagem do problema, levando a uma solução rápida. Como consequência, decisões mais rápidas e aparentemente melhores serão tomadas.

Uma atividade de negócios precisa ser inteligentemente automatizada para ser monitorada. O monitoramento deve ser inteligente, e os resultados devem ser fáceis de acessar, visualizar ou agir para produzir valor. Modelagem da atividade é a primeira etapa na criação de um sistema BAM bem-sucedido. Envolve encontrar atividades que devam ser monitoradas, definir suas etapas e eventos, e ligar esses eventos às métricas de desempenho que serão monitoradas (veja McKie, 2003). Atividades de análise resultam em processos aprimorados.

Questões de BAM

Sistemas BAM sofrem com as muitas deficiências de todos os outros EIS. Com frequência, executivos falham ao levar em consideração a prestação da tecnologia ou dos processos de negócios que eles querem monitorar. Essa era uma das razões para falha no sistema empresarial da McDonald's. Sistemas empresariais requerem um líder sênior de gerenciamento e envolvimento e treinamento adequados dos usuários. Questões de integração podem incomodar uma iniciativa de BAM. Os dados devem ser extraídos de muitas fontes diferentes e fornecidos a muitos usuários.

Questões de gerenciamento de mudanças são as mais importantes. Assim como com a maioria dos EIS, adotar modelo e método exige não apenas tecnologia, mas também uma mudança nos processos de negócios (ou seja, gerenciamento de mudanças). No caso do BAM, são eliminados os atrasos incorporados aos fluxos de informação para permitir que gerentes de nível mais baixo os respondam. Isso tem causado problemas no EIS, e gerentes de alto nível devem aprender a deixar alguns problemas nas mãos desses responsáveis, pelo menos por um tempo, para lhes possibilitar a oportunidade de responder. Executivos não devem enfraquecer a autonomia de seus funcionários.

Assim como com qualquer EIS, o BAM efetivo requer funcionamento estreito com as unidades de negócios a fim de identificar os indicadores-chave (FCS) e as técnicas analíticas que fornecem advertências preliminares confiáveis de questões iminentes (ou seja, alertas). Também, como para qualquer outro EIS, uma boa maneira de iniciar um BAM em tempo real é se concentrar em um problema de negócio bem-definido com um retorno demonstrável (veja Keating, 2003).

BAM tem a capacidade de dar a informação certa às pessoas que precisam dela de forma mais rápida. Simultaneamente, a informação é relatada aos níveis mais altos da gerência. Os executivos devem deixar os gerentes responsáveis nas linhas de frente enfrentarem os próprios problemas e questões em tempo hábil antes de reagir em (veja a próxima seção). A chave para o sucesso é fornecer àqueles mais próximos à situação as informações que precisam para tomada de decisão e, ao mesmo tempo, ajudar os níveis superiores de gerenciamento a monitorar de modo mais efetivo as consequências das decisões (veja Keating, 2003).

Fornecedores de BAM

O mercado de *software* de BAM está nos estágios iniciais de desenvolvimento. Como qualquer outro mercado de *software* emergente, está experimentando problemas de definição e crescimento. Alguns anos atrás, existiam uma série de fornecedores especializados em *software* de BAM. Hoje, é difícil encontrar um fornecedor que se nomeie como estritamente de BAM. A Synthean e a Quantive são duas exceções de empresas especializadas. Algumas das antigas empresas especializadas, como a SeeRun, não existem mais ou, como a Elity, foram compradas. Entretanto, muitas focalizaram em serviços verticais especiais (p.ex., FirstRain) ou refocalizaram em aspectos particulares de gerenciamento de desempenho (p.ex., a mudança da Celequest para gerenciamento de desempenho operacional e *dashboards* operacionais). Por fim, diversas empresas de *software* oferecem componentes ou funcionalidades de BAM como parte dos seus amplos pacotes de gerenciamento de desempenho ou BI. Cognos, Tibco, Savvion e Mentisys são alguns dos fornecedores que dão suporte a BAM nas suas ofertas abrangentes. Para uma lista detalhada, mas levemente obsoleta, de fornecedores de BAM, acesse computerworld.com/softwaretopics/software/story/0,10801,86894,00.

Questões de revisão da Seção 5.11

1. Defina *BAM*.
2. Descreva o sistema BAM usado pelo governo da cidade de Albuquerque.
3. Quais são os principais elementos técnicos de um sistema BAM?
4. Quais são algumas das principais questões que surgem com a implementação de BAM?

Destaques do capítulo

- BPM é um termo guarda-chuva que abrange metodologias, métricas, processos e sistemas de negócios usados para promover o desempenho total de uma empresa.
- BPM é um desdobramento da BI e incorpora muitas de suas tecnologias, aplicações e técnicas.
- BPM inclui um conjunto de processos de ciclo fechado que liga estratégia à execução a fim de otimizar o desempenho de negócios.
- Os principais processos de BPM são: definir estratégias, planejar, monitorar, agir e ajustar.
- Atualmente, as empresas reconhecem o valor do planejamento estratégico, bem como a necessidade de excelência operacional e em execução.
- Estratégia responde a questão “Aonde queremos ir no futuro?”
- A lacuna entre estratégia e execução é encontrada nas extensas áreas de visão, pessoas, gerenciamento e recursos.
- Planos operacionais e táticos enfocam a questão “Como chegamos ao futuro?”
- As táticas e iniciativas definidas em um plano operacional precisam ser diretamente vinculadas aos objetivos e às metas principais no plano estratégico.
- Os objetivos estratégicos e as métricas principais de uma empresa devem funcionar como propulsores de cima para baixo para a alocação dos ativos tangíveis e intangíveis de uma empresa.
- Monitoramento enfoca a questão de “Como estamos fazendo?”
- BSC, *dashboards* de desempenho, sistemas de monitoramento de projeto, sistemas de recursos humanos e sistemas de relatório financeiro são todos exemplos de sistemas de controle de diagnóstico.
- A maioria dos monitoramentos concentra-se nas variâncias negativas e presta pouca atenção nas premissas ou estratégias subjacentes.

- Existe uma diferença entre planejamento para uma linha de negócios convencional e planejamento para um novo empreendimento.
- Segundo estudos de marca comparativa, a empresa de bilhões de dólares típica dedica milhares de pessoas-dia a cada ano para planejar e administrar uma geração de relatório.
- Números de pesquisa indicam que a alta gerência gasta cerca de três horas por mês discutindo estratégias e 80% do seu tempo em questões que respondem por menos de 20% do valor a longo prazo da empresa.
- Em empresas que têm as melhores práticas, previsão é um processo de suporte à decisão que é usado para refinar o raciocínio estratégico, revisar planos e ajustar alocação de recursos em resposta aos eventos em constante mudança.
- Sistemas de medição de desempenho auxiliam os gerentes no controle da implementação da estratégia de negócios através da comparação entre resultados atuais e metas e objetivos estratégicos.
- As desvantagens de usar dados financeiros como o núcleo de um sistema de medição de desempenho são bem-conhecidas.
- Medidas de desempenho precisam ser derivadas de estratégias corporativas e da unidade de negócios e de uma análise dos principais processos de negócios exigidos para realizar essas estratégias.
- Provavelmente, o sistema de gerenciamento de desempenho mais bem conhecido e mais amplamente usado é o BSC.
- Central à metodologia de BSC está uma visão holística do sistema de medida ligada à orientação estratégica da empresa.
- Como uma metodologia de medida, o BSC é planejado para superar as limitações de sistemas que têm foco financeiro.
- Como uma metodologia de gerenciamento estratégico, o BSC possibilita que uma empresa alinhe suas ações com suas estratégias gerais.
- A maioria das empresas usa Six Sigma como metodologia para melhoria de processo, o que lhes permite esmiuçar seus processos, identificar problemas e aplicar soluções.
- Six Sigma é uma metodologia de gerenciamento de desempenho com objetivo de reduzir o número de falhas em um processo de negócio para o mais próximo de zero DPMO possível.
- Six Sigma usa DMAIC, modelo de melhoria de negócios de circuito fechado que inclui as etapas de definição, medição, análise, melhoria e controle de um processo.
- Enormes benefícios de desempenho podem ser obtidos com a integração entre BSC e Six Sigma.
- A arquitetura de BPM padrão é multicamada e consiste em uma camada de interface do usuário, uma camada de banco de dados e uma camada de dados fonte.
- As principais aplicações de BPM incluem orçamento, planejamento e previsão; análise e otimização de lucratividade; *scorecarding*; consolidação financeira; e geração de relatórios financeiros e legais.
- Indicadores e *dashboards* são componentes comuns da maioria, se não de todos, os sistemas de gerenciamento de desempenho e pacotes de BPM.
- Embora os indicadores e *dashboards* forneçam exibições visuais do desempenho, existem diferenças significativas entre eles.
- O desafio essencial da estrutura de *dashboard* é mostrar todas as informações necessárias em uma única tela, nitidamente e sem distração, de maneira que possam ser rapidamente assimiladas.
- BAM é usado para supervisionar atividades em tempo real ou eventos de uma determinada instalação, como uma fábrica ou central de atendimento, ou um processo de negócios específico.
- BAM ajuda não apenas no reconhecimento e na resposta a eventos, mas também na possibilidade de gerentes resolverem ocorrências de evento rapidamente e analisarem suas influências para tomarem decisões mais informadas e em tempo hábil.

Termos-chave

- | | | |
|--|---|--------------------------------------|
| • arquitetura de sistema | • fator crítico de sucesso (FCS) | • objetivo estratégico |
| • <i>balanced scorecard</i> (BSC) | • indicador de desempenho | • plano operacional |
| • <i>business performance management</i> (BPM) | • mapa estratégico | • sistema de controle de diagnóstico |
| • <i>dashboard</i> de desempenho | • meta estratégica | • sistema de medida de desempenho |
| • DMAIC | • monitoramento de atividades de negócios (BAM) | • Six Sigma |
| | | • visão estratégica |

QUESTÕES PARA DISCUSSÃO

1. A SAP usa o termo gestão estratégica de empresas (SEM), a Cognos usa o termo gerenciamento de desempenho corporativo (CPM) e a Hyperion usa o termo gerenciamento do desempenho do negócio (BPM). Elas estão referindo-se às mesmas idéias básicas? Forneça evidências para comprovar sua resposta.
2. BPM inclui quatro processos básicos: definir estratégia, planejar, monitorar, agir e ajustar. Selecione um desses processos e discuta os tipos de ferramentas e aplicações de *software* disponíveis para suportá-lo.
3. Selecione uma empresa pública interessante. Usando o relatório anual de 2006 da empresa, crie cinco objetivos estratégicos financeiros para 2007. Para cada objetivo, especifique uma meta estratégica. As metas devem ser compatíveis com o desempenho financeiro da empresa em 2006.
4. Nos últimos anos, o Beyond Budgeting Round Table (BBRT; bbirt.org) questionou práticas tradicionais de orçamento. Uma série de artigos na Web discute a posição do BBRT. Na visão do BBRT, o que está errado nas práticas de orçamento de hoje? O que o BBRT recomenda como substituto?
5. Descreva como um BSC se encaixa na descrição de um sistema de controle de diagnóstico.
6. Distinga gerenciamento de desempenho de medida de desempenho.
7. O Prisma de Desempenho, criado por Andrew Neely e outros na Universidade de Cranfield, fornece um modelo alternativo de gerenciamento e medida de desempenho. Usando materiais da Web, discuta os principais pressupostos do modelo. Compare-o ao BSC e Six Sigma.
8. Crie uma medida para qualquer objetivo estratégico interessante (você pode usar um dos objetivos formulados na Questão de discussão 3). Para a medida selecionada, complete o modelo de medida localizado na Tabela W5.3.1 do arquivo *online* para este capítulo.
9. Crie uma estratégia para uma empresa hipotética, usando as quatro perspectivas do BSC. Represente-a como uma série de objetivos estratégicos. Produza um mapa estratégico que descreva os vínculos entre os objetivos.
10. Compare e contraste o modelo DMAIC com os processos de BPM de ciclo fechado.
11. O quadrante mágico de BPM do Gartner lista os principais fornecedores de *software* que oferecem pacotes de BPM. Compare e contraste duas das ofertas (exceto o pacote Hyperion Solutions).
12. Diversas empresas fornecem tanto produtos de indicadores quanto de *dashboards*. Compare e contraste as ofertas de indicadores e *dashboards* de uma empresa.
13. Usando o critério determinado na Tabela 5.2, compare BAM com BI tradicional e BI para BPM. Com base nessa comparação, BAM é mais semelhante a qual tipo de BI?

Exercícios

Exercícios da Teradata University e outros exercícios interativos

1. Acesse teradatastudentnetwork.com e procure por “*performance management*” digitando essa frase na caixa Search. Após os resultados terem sido fornecidos, selecione o seminário na Web “Facilitating High-Performance Business”. Esse seminário descreve o uso dos recursos de BI e gerenciamento de desempenho da Cognos pela Harrah’s. Com base nessa discussão, responda as seguintes questões:
 - a. Quais pressões estão impulsionando o crescimento da CPM?
 - b. Qual é o valor estratégico da CPM?
 - c. Quais são os elementos centrais da CPM?
 - d. Quais são as três principais áreas de aplicação da CPM?
 - e. Quais eram as questões estratégicas (pontos fracos) que motivaram as mudanças nos recursos analíticos e de gerenciamento de desempenho da Harrah’s?
 - f. Faça uma breve descrição do sistema de *marketing* de ciclo fechado da Harrah’s.
 - g. Quais tecnologias são usadas no sistema de *marketing* da Harrah’s?
 - h. Como a análise de ciclo fechado da Harrah’s assemelha-se ao ciclo fechado para BPM?
 - i. Quais KPI são de interesse básico para a Harrah’s?
 - j. Quais lições a Harrah’s aprendeu sobre implementação e uso do seu sistema?
2. Acesse teradatastudentnetwork.com. Selecione os *demos* na guia Resources. Assista o demo “Supply Chain Intelligence (SCI)”. Faça um breve resumo do cenário demonstrado. Descreva os recursos de BAM fornecidos pela SCI e análise preditiva da Teradata.

3. Existem vários artigos que descrevem as armadilhas associadas à estrutura de *dashboard*. Um desses artigos é “*Dashboard Design: Why Design is Important*”, de Richard Brath e Micheal Peters. Faça o *download* (dmreview.com/article_sub.cfm?articleId=1011285) e leia. Em seguida, acesse dundas.com. A Dundas fornece componentes para desenvolvimento de aplicações personalizadas de indicadores. No menu “Product”, selecione os componentes “*gauges*”.NET. Então, vá até os *demos* de medida. Execute o *demo* Executive Digital *Dashboard*. Navegue pelas diversas guias e responda as seguintes questões:
 - a. Quais tipos de informações são fornecidas pelo *dashboard* executivo? Faça uma breve descrição do status da empresa representado por esse *demo*.
 - b. Usando os conceitos do artigo de Brath e Peters, faça uma crítica da estrutura de *dashboard* do *demo*. De que maneiras a estrutura poderia ser melhorada?
4. Desenvolva um protótipo de *dashboard* para mostrar os resultados financeiros de uma empresa pública. O protótipo pode ser feito em papel ou no Excel. Use os dados dos planos anuais de 2006 de duas empresas públicas para ilustrar os recursos do seu *dashboard*.

Tarefas em grupo e interpretação de papéis

1. Em grupo, selecione um fornecedor específico do quadrante mágico de BPM (exceto Hyperion). Para cada fornecedor, resuma as ofertas de BPM, a arquitetura e as aplicações fornecidas. Praticamente todos os fornecedores oferecem estudos de caso que descrevem as implementações de BI e BPM dos seus clientes. Selecione três estudos de caso. Para cada um, resuma o problema que o cliente estava tentando enfrentar, as aplicações ou soluções implementadas e os benefícios recebidos pelo cliente a partir do sistema.
2. A Computerworld oferece uma lista razoavelmente completa de fornecedores de BAM (acesse computerworld.com/software-topics/software/story/0,10801,86894,00). Cada grupo deve selecionar um fornecedor específico dessa lista. Para cada fornecedor, resuma as ofertas de BAM, a arquitetura e as aplicações fornecidas. Praticamente todos os fornecedores oferecem estudos de caso que descrevem as implementações de BAM dos seus clientes. Selecione três estudos de caso. Para cada um, resuma o problema que o cliente estava tentando enfrentar, as aplicações ou soluções implementadas e os benefícios recebidos pelo cliente a partir do sistema.

Exercícios na Internet

1. Uma pesquisa realizada pela Economist Intelligence Unit e apresentada por S. Taub, “Closing the Strategy-to-Performance Gap”, *CFO Magazine*, 22 de fevereiro de 2005 (cfo.com/article.cfm/3686974?f=related), explora a relação entre desenvolvimento e execução de estratégia. Com base nessa pesquisa, o que é mais importante para o gerenciamento de desempenho – desenvolvimento ou execução da estratégia? Quais são as razões dadas pelos entrevistados para uma execução ineficiente? De que maneira os entrevistados pensam que podem melhorar o desempenho?
2. No seu artigo “Discover-Driven Planning”, McGrath e MacMillan propõem um método para planejamento de novos empreendimentos diante de premissas incertas. Um artigo recente de Sunil Sharma, “Discovery Driven Planning”, aplicou o método a uma pequena empresa fictícia (para consistência, marketremarks.com/narad.php?ContentId=10). O artigo também fornece acesso à planilha usada no exemplo. Usando essa planilha, desenvolva um *dashboard* que a empresa fictícia poderia usar para controlar seus planos.
3. Um artigo de Tom Hoblitzell, “Best Practices in Planning and Budgeting”, resume muitas das descobertas do estudo de marca comparativa do The Hackett Group. Faça o *download* desse artigo em answerthink.com/pdf/pdf_services/04_intelligence/whitepapers/wp_bpm_planbudget.pdf. Escreva um relatório resumindo as principais descobertas desse estudo.
4. A BSC Collaborative (bscol.com) oferece uma lista de fornecedores que vendem produtos de BSC certificados (bscol.com/bsc_online/technology/). Compare e contraste dois dos produtos.
5. Recentemente, a Tektronix ganhou o CMO Best Practices Award por seu *dashboard* de *marketing*. O sistema está descrito em BtoBOnline.com, (bto-bonline.com/article.cms?articleId=23849). Faça uma breve descrição do sistema. Quais são algumas das métricas controladas pelo sistema? Quais benefícios a Tektronix obtém do sistema?
6. A Hyperion Solutions fornece um documento informativo intitulado “Business Performance Management” (hyperion.com/downloads/wp_bpm_0603.pdf) que descreve o modelo de gerenciamento da Hyperion. Faça o *download* desse documento. Descreva o modelo de gerenciamento da Hyperion. Compare-o com os processos de BPM de ciclo fechado descritos neste capítulo.
7. Um artigo recente intitulado “Is BPM Performing? Surveys Highlight the Market’s Needs”, de Meg

Waters (acesse bpmmag.net/magazine/article.html?articleID=14437), discute os resultados de duas pesquisas de BPM distintas projetadas para responder a questão geral “Quão bem os fornece-

dores de BPM estão distribuindo?”. Descreva as pesquisas. Faça um resumo dos resultados das pesquisas discutidos no artigo.

CASO DE APLICAÇÃO DE FINAL DO CAPÍTULO

Sistemas de informação vigilantes na Western Digital

A Western Digital (WD) é uma criadora e fabricante global de discos rígidos de alto desempenho para *desktops*, redes corporativas, armazenamento empresarial e aplicação de entretenimento doméstico que vale US\$3 bilhões. A empresa está localizada em Lake Forest, Califórnia, e possui fábricas na Malásia e Tailândia e centros de distribuição na Europa. Ela emprega cerca de 18.000 pessoas no mundo todo.

Durante a última década, a competição no setor de discos rígidos foi feroz. Como muitos outros setores, ele também enfrenta constantes mudanças nas exigências do cliente, pressões de preço devido à competição global, curta vida útil do produto e alta demanda de qualidade e confiabilidade do produto. Como resultado, o setor recuou de 11 fabricantes para 3 a 5, dependendo da linha de produto. Nesse período, a WD se sobressaiu e, agora, é a terceira maior fabricante em volume.

NECESSIDADE DE NOVO FORNECIMENTO DE INFORMAÇÃO

Como parte da sua estratégia de sobrevivência, o gerenciamento da WD manifestou uma necessidade de uma nova maneira de fornecer informação. Eles queriam a capacidade de reagir mais rapidamente. Também queriam informações integradas para que pudessem gerenciar toda a empresa da maneira “seguindo o sol”, transmitindo informações em todos os fusos horários à medida que o dia de trabalho se encerra em um e começa em outro. A solução para essas necessidades é o Sistema de Informação Vigilante (VIS) da WD.

A palavra *vigilante* significa estar sempre atento. O VIS “integra dados e informações puras de diversas fontes a fim detectar mudanças, criar alertas, auxiliar na diagnose de análise de problemas, e dar suporte à comunicação para uma ação rápida.” Diferente dos sistemas de informação tradicionais, o VIS é proativo ao invés de passivo. Os dados são reanalisados à medida que as mudanças são feitas neles. Se os dados cumprem certas condições preestabelecidas, o sistema alerta o usuário.

ARQUITETURA DO VIS

O VIS da WD tem uma arquitetura de multicamada. Na parte inferior do sistema estão os dados brutos provenientes de diversas fontes de operação. Esses dados vão para uma série de aplicações funcionais (p.ex., ERP, pontos de venda, fabricação). Na parte superior dessa camada está a camada de BI, que analisa os dados para determinar se eles cumprem certas condições preestabelecidas. Os dados que satisfazem essas condições iniciam os alertas que são enviados à camada de *dashboard* na parte superior.

Três recursos formam a base do VIS da WD: o sistema ERP, que foi implantado em 1997 e fornece dados atualizados sobre as operações da empresa; o *data warehouse*, que foi implantado em 1999 e fornece dados integrados de 12 sistemas legados distintos; e o Sistema de Informação de Qualidade (QIS), que foi implantado em 1999 e fornece *insights* na qualidade geral de fabricação. Embora esses sistemas formem a base do VIS e tenham estado em vigor por algum tempo, eles falharam no fornecimento de visão das operações necessárias para os gerentes. Em particular, alguns sistemas legados ainda continuaram não integrados, a taxa de atualização de dados dos sistemas era inadequada, e os sistemas careciam de recursos analíticos adequados.

A gerência percebeu que o VIS combinado com os *dashboards* de gerenciamento em tempo real poderia fornecer a visibilidade desejada, desde que mudanças em determinadas regras fossem instituídas. Primeiro, eles tiveram de traduzir as metas estratégicas da empresa em um conjunto de objetivos operacionais alinhados, mensuráveis e baseados em tempo para cada departamento. Em seguida, tiveram de monitorar os KPIs em tempo real, tanto horizontalmente ao longo da empresa quanto verticalmente dentro das unidades de negócios. Por fim, tiveram de estimular tomadas de decisão colaborativas entre grupos em equipes, departamentos, empresas e regiões geográficas.

DASHBOARDS

Dois tipos de *dashboards* em tempo real foram desenvolvidos: um para dados da fábrica e outro para

informações corporativas, incluindo planejamento de demanda, distribuição e vendas. Os *dashboards* de fábrica – projetados para produção, material, saída de produto, monitoramento de estação e qualidade – foram criados do zero. Cada um dos cinco *dashboards* de fábrica mostrava diversos KPI e métricas, permitia *drill-down* em detalhes subjacentes e emitia alertas. Por outro lado, os *dashboards* corporativos – projetados para faturamento e lucro, pedidos pendentes, perspectiva, estoque de bens acabados e afins – foram criados usando o pacote de BI da Cognos. O pacote forneceu os mesmos recursos que os *dashboards* de fábrica, bem como uma série de outros recursos analíticos especializados.

Além de fornecer a visibilidade necessária aos sistemas operacionais da WD, a combinação entre VIS e *dashboards* também deu suporte à metodologia de decisão OODA da WD. OODA significa processos integrados de *observação* (ou seja, ver sinais de mudança), *orientação* (ou seja, interpretar os sinais), *decisão* (ou seja, formular uma resposta adequada) e *ação* (ou seja, executar a resposta selecionada). A partir da perspectiva da WD, as empresas que conseguem executar ciclos OODA mais rápido têm uma vantagem competitiva.

IMPACTO NOS NEGÓCIOS

Em 2004, mais de 200 gerentes e profissionais de todos os níveis na WD usavam os *dashboards*. Os benefícios provenientes do sistema e dos *dashboards* vieram de duas formas:

- **Redução de custos.** O sistema e os *dashboards* forneceram uma melhor visão dos giros de estoque e de outros custos de estocagem. O resultado foi uma economia líquida de US\$3 milhões. O sistema também reduziu os custos de produção de relatórios de gerenciamento personalizados. Aqui, a economia é estimada em US\$900.000 por ano. Por fim, o sistema reduziu a sobrecarga de informação e, por consequência, a quantidade de tempo gasto em reuniões. Nesse caso, a redução está estimada em US\$350.000 por ano.
- **Vantagens estratégicas.** O sistema e os *dashboards* aceleram o fornecimento de dados e informações aos executivos e gerentes de todos os níveis. Isso significa que problemas e oportunidades podem ser tratados mais rapidamente, que decisões estratégicas podem ser tomadas mais prontamente e que os processos no ciclo OODA podem ser executados mais rapidamente.

LIÇÕES APRENDIDAS

As experiências da WD fornecem várias lições para implantação e uso de VIS e *dashboards* de desempenho. Algumas das lições mais importantes incluem:

- Projetar os *dashboards* de gerenciamento em tempo real para que sejam a espinha dorsal da administração da empresa.
- Planejar e programar a coordenação entre equipes, de modo a usar os *dashboards* para gerenciar toda a empresa.
- Criar um ciclo de aprendizagem sobre cada ciclo OODA a fim de estimular o aprendizado do grupo, pois quanto mais rápido o ciclo, mais importantes as revisões de aprendizado e mais frequentes elas precisam ser.
- Combinar a latência de cada ciclo OODA com as necessidades e habilidades da empresa para tornar-se verdadeiramente vigilante.
- Fornecer os blocos de construção para a empresa “perceber e responder” em tempo real através do VIS e dos *dashboards* de gerenciamento em tempo real.
- Justificar o VIS sobre uma base estratégica em vez de sobre um custo ou base de retorno do investimento.
- Tornar a implantação de um VIS em toda a empresa uma iniciativa de gerenciamento porque ela exige comprometimento colaborativo dinâmico da alta gerência para introduzir a transformação organizacional necessária.

Fontes: compilado de HOUGHTON, R., EL SAWY, O.A., GRAY, P., DONEGAN, C. e JOSHI, A. Vigilant Information System for Managing Enterprises in Dynamic Supply Chains: Real-Time *Dashboards* at Western Digital. *MIS Quarterly*, março de 2004; western.digital.com (acessado em outubro de 2006).

QUESTÕES SOBRE O CASO

1. O que é VIS?
2. De que maneira a arquitetura do VIS é semelhante ou diferente da arquitetura de BPM?
3. Quais são as semelhanças e diferenças entre os processos de BPM de ciclo fechado e os processos no ciclo de decisão OODA?
4. Usando os tipos de *dashboards* de Eckerson, descreva os tipos de *dashboards* oferecidos pelo sistema da WD.
5. Quais são os benefícios básicos fornecidos pelo VIS e pelos *dashboards* da WD?
6. Quais tipos de conselho você pode fornecer a uma empresa que está se preparando para criar seus próprios VIS e *dashboards*?

Referências

- Axson, D. (2003). *Best Practices in Planning and Management Reporting*. Hoboken, NJ: Wiley.
- BPM Standards Group. (2005, September). *Business Performance Management: Industry Framework Document*, bpmstandardsgroup.org/documents/BPMIndustryFramework-V5.0090905.pdf (acessado em março de 2006).
- BSC Collaborative. (2000). *Functional Standards*, bscol.com/pdf/bsc99-standards.pdf (acessado em março de 2006).
- Buytendijk, F., et al. (2004, January). "Drivers and Challenges of Corporate Performance Management". *Gartner Research Report R-22-0730*.
- Charan, R., and G. Colvin. (1999, July). "Why CEOs Fail". *Fortune*.
- Coveney, M., et al. (2003). *The Strategy Gap*. Hoboken, NJ: Wiley.
- Eagle, J. (2005, March). *Cross-Industry Study by Economist Intelligence Unit Finds More Companies May Undertake Six Sigma This Year*, sixsigmazone.com/press-releases/cross-industry-study-by-economistintelligence-unit (acessado em março de 2006).
- Eckerson, W. (2006). *Performance Dashboards*. Hoboken, NJ: Wiley.
- Few, S. (2005, Winter). "Dashboard Design: Beyond Meters, Gauges, and Traffic Lights". *Business Intelligence Journal*.
- Gupta, P. (2004). *Six Sigma Business Scorecard*. New York: McGraw-Hill.
- Hackett Group. (2004, August). *Gap Between World-Class and Median Continues to Widen According to 2004 Performance Research from The Hackett Group*, answerthink.com/04_news/01_press/pr_2004/pr_08102004_01.html (acessado em março de 2006).
- Hammer, M. (2003). *Agenda: What Every Business Must Do to Dominate the Decade*. Pittsburgh, PA: Three Rivers Press.
- Hoblitzell, T. (2003). *Best Practices in Planning and Budgeting*, answerthink.com/pdf/pdf_services/04_intelligence/whitepapers/wp_bpm_planbudget.pdf (acessado em março de 2006).
- Isixsigma. (2006). *The History of Six Sigma*, 2006, isixsigma.com/library/content/c020815a.asp (acessado em março de 2006).
- Kaplan, R., and D. Norton. (1992, January–February). "The Balanced Scorecard – Measures That Drive Performance". *Harvard Business Review*, pp. 71–79.
- Kaplan, R., and D. Norton. (1996). *The Balanced Scorecard: Translating Strategy into Action*. Boston: Harvard University Press.
- Kaplan, R., and D. Norton. (2000). *The Strategy-Focused Organization: How Balanced Scorecard Companies Thrive in the New Business Environment*. Boston: Harvard Business School Press.
- Kaplan, R., and D. Norton. (2004). *Strategy Maps: Converting Intangible Assets into Tangible Outcomes*. Boston: Harvard Business School Press.
- Keating, W. (2003, March). "FastTracking", optimize-mag.com/article/showArticle.jhtml?articleId=17700874 (acessado em março de 2006).
- McGrath, R., and I. MacMillan. (1995, July–August). "Discovery-Driven Planning". *Harvard Business Review*.
- McKie, S. (2003, July 18). "The Big BAM". *Intelligent Enterprise*, intelligententerprise.com/030718/612feat3_1.jhtml?jsessionid=KSXCENUTF1TDYQSNL0SKH0CJUNN2JVN (acessado em outubro de 2006).
- A. Neely, A., and M. Bourne. (2000, December). "Why Measurement Initiatives Fail". *Measuring Business Excellence*, Vol. 4, No. 1.
- Niven, P. (2005). *Balanced Scorecard Diagnostics*. Hoboken, NJ: Wiley.
- Porter, M.E. (2001, March). "Strategy and the Internet". *Harvard Business Review*.
- Rayner, N., et al. (2005a, October). "Corporate Performance Management Applications Explained". *Gartner RAS Core Research Note G00130282*.
- Rayner, N., et al. (2005b, October). "Magic Quadrant for CPM Suites, 2005". *Gartner Research Note G00131208*.
- Schiemann, W., and Lingle, J. (1999). *BULLSEYE! Hitting Your Strategic Targets Through High-Impact Measurement*. New York: The Free Press.
- Simons, R. (2002). *Performance Measurement and Control Systems for Implementing Strategy*. Upper Saddle River, NJ: Prentice Hall.
- Systar. (2002, April). *Business Activity Monitoring*, systar.com/docs/solutions_bam_wp.pdf (acessado em abril de 2006).
- Tarlow, M., and Tarlow, P. (2002). *Digital Aboriginal*. New York: Warner.
- Wade, D., and R. Recardo. (2001). *Corporate Performance Management*. Boston: Butterworth-Heinemann.
- Whiting, R. (2005, April). "A Closer Look at Performance". *InformationWeek*, informationweek.com/story/showArticle.jhtml?articleID=160901247 (acessado em abril de 2006).
- Wikipedia.org. (2006, January). *Strategic Enterprise Management*, en.wikipedia.org/wiki/Strategic_enterprise_management (acessado em abril de 2006).

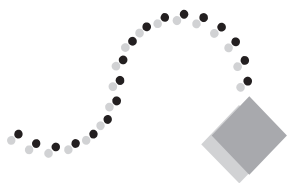


Redes Neurais para *Data Mining*

Objetivos de aprendizado

- ◆ Compreender os diferentes conceitos e tipos de redes neurais artificiais (ANN)
- ◆ Estudar as vantagens e limitações das ANN
- ◆ Compreender como a propagação de retaguarda de uma rede neural funciona
- ◆ Compreender o processo completo de uso de redes neurais
- ◆ Apreciar a completa variedade de aplicações de redes neurais

Este capítulo está disponível *online* no endereço www.bookman.com.br.



Glossário

- administrador do data warehouse (DWA)** pessoa responsável pela administração e gerenciamento de um *data warehouse*.
- agrupamento** divisão de um banco de dados em segmentos nos quais os membros de um segmento compartilham qualidades semelhantes.
- algoritmo de aprendizado** procedimento de treinamento usado por uma rede neural artificial.
- análise da Web** aplicação de atividades analíticas de negócios a processos baseados na Web, incluindo comércio eletrônico.
- análise de acesso na Web** análise de dados que ocorre em um ambiente Web.
- análise de negócios (BA)** aplicação de modelos diretamente aos dados comerciais. BA envolve o uso de ferramentas de MSS, especialmente modelos, no auxílio aos tomadores de decisão. Basicamente, é uma forma de suporte à decisão OLAP.
- análise preditiva** uso de ferramentas que ajudam a determinar o provável resultado futuro para um evento ou a probabilidade de uma situação ocorrer. Essas ferramentas também identificam relações e padrões.
- analítica** ciência da análise.
- aprendizado não supervisionado** método de treinamento de redes neurais artificiais no qual somente estímulos de entrada são mostrados às redes, que é auto-organizável.
- aprendizado supervisionado** método de treinamento de redes neurais artificiais no qual casos de exemplo são mostrados à rede como entrada e os pesos são ajustados para minimizar o erro das saídas.
- arquitetura de sistema** estruturas lógica e física de um sistema.
- árvore de decisão** técnica para classificação de entidades específicas em classes singulares com base nas características das entidades. Uma raiz é seguida por nós internos. Cada nó (incluindo a raiz) é rotulado com uma questão, e arcos associados a cada nó abrangem todas as respostas possíveis.
- associação** categoria de algoritmo de *data mining* que estabelece relações a respeito de itens que ocorrem juntos em um determinado registro.
- auto-organização** arquitetura de rede neural que usa aprendizado não supervisionado.
- axônio** conexão de saída (ou seja, final) de um neurônio biológico.
- balanced scorecard (BSC)** metodologia de gerenciamento e medição de desempenho que ajuda a traduzir processos financeiros, de cliente e internos, e objetivos e metas de aprendizado e crescimento da empresa em um conjunto de iniciativas acionáveis.
- banco de dados multidimensional** banco de dados onde os dados são organizados especificamente para suportar análise multidimensional fácil e rápida.
- business performance management (BPM)** abordagem avançada de medição e análise de desempenho que inclui planejamento e estratégia.
- camada escondida** camada central de uma rede neural artificial que tem três ou mais camadas.
- classificação** indução supervisionada usada para analisar os dados históricos armazenados em um banco de dados e gerar automaticamente um modelo que possa prever comportamento futuro.
- computação neural** projeto de computação experimental com objetivo de criar computadores inteligentes que operem de maneira semelhante ao funcionamento do cérebro humano. *Ver também* rede neural artificial (RNA).
- consulta ad hoc** consulta que não pode ser determinada antes do momento em que a consulta é realizada.
- controle de BI** processo de priorização de projeto dentro das empresas.
- cubo de dados** objeto bidimensional, tridimensional ou com uma dimensão superior no qual cada dimensão de dados representa uma medida de interesse.
- cubo** subconjunto de dados altamente inter-relacionados organizado para permitir aos usuários combinar quaisquer atributos em um cubo (p.ex., lojas, produtos, clientes, fornecedores) com quaisquer métricas no cubo (p.ex., vendas, lucro, unidades, idade) a fim de criar diversas visões bidimensionais.

nais, ou visões, que podem ser exibidas em uma tela de computador.

dados de acesso na Web dados que fornecem uma pista das atividades do usuário e mostram os padrões de navegação do usuário (p.ex., quais *sites* são visitados, quais páginas, por quanto tempo).

dashboard de desempenho exibição visual usada para monitorar desempenho operacional.

dashboard apresentação visual de dados críticos para executivos. Permite aos executivos enxergar “*pon-tos quentes*” em segundos e explorar a situação.

data mart dependente subconjunto que é criado diretamente de um *data warehouse*.

data mart independente *data warehouse* pequeno projetado para uma unidade estratégica de negócios ou um departamento.

data mart *data warehouse* departamental que armazena apenas dados relevantes.

data mining baseado em descoberta forma de *data mining* que encontra padrões, associações e relações entre dados a fim de descobrir fatos previamente desconhecidos ou nem mesmo considerados pela empresa.

data mining baseado em hipótese forma de *data mining* que começa com uma proposição pelo usuário, que então busca validar a veracidade da proposição.

data mining processo que usa técnicas estatísticas, matemáticas, de inteligência artificial e de aprendizado de máquina para extrair e identificar informações úteis e conhecimento subsequente de extensos bancos de dados.

data warehouse empresarial (EDW) *data warehouse* de nível organizacional desenvolvido para fins analíticos.

data warehouse depósito físico onde dados relacionais são especialmente organizados para fornecer dados limpos em um formato padronizado para toda a empresa.

data warehousing ativo *ver data warehousing* em tempo real.

data warehousing em tempo real processo de carregar e fornecer dados através de um *data warehouse* à medida que eles se tornam disponíveis.

dendrito parte de um neurônio biológico que fornece estímulos à célula.

descoberta de sequência identificação de associações ao longo do tempo.

DMAIC modelo de melhoria de negócios de circuito fechado que inclui as seguintes etapas: definir, medir, analisar, melhorar e controlar um processo.

drill-down investigação de informações em detalhes (p.ex., encontrar não apenas total de vendas, mas também vendas por região, por produto ou por vendedor). Encontrar as fontes detalhadas.

elemento de processamento (PE) neurônio de uma rede neural.

entropia métrica que avalia o grau de incerteza ou aleatoriedade em um conjunto de dados. Se todos os dados em um subconjunto pertencerem a apenas uma classe, então não há incerteza ou aleatoriedade naquele conjunto de dados e, portanto, a entropia é zero.

extração, transformação e carga (ETL) processo de *data warehousing* que consiste em extração (ou seja, leitura de dados de um banco de dados), transformação (ou seja, conversão dos dados extraídos de sua forma anterior na forma em que precisam estar, para que sejam colocados em um *data warehouse* ou apenas em outro banco de dados) e carga (ou seja, colocação dos dados no *data warehouse*).

fatores críticos de sucesso (FCS) fatores-chave que descrevem as ações que uma empresa deve primar para ser bem-sucedida no seu espaço de mercado.

função de soma mecanismo para adicionar todas as entradas vindas de um determinado neurônio.

função de transformação (transferência) em uma rede neural, a função que soma e transforma entradas antes da descarga do neurônio. Mostra a relação entre o nível de ativação interno e a saída de um neurônio.

função sigmóide (ativação lógica) função de transferência em forma de S na variação de 0 a 1.

gerenciamento de decisões empresariais (EDM) *ver* suporte automatizado à decisão.

grão definição do nível mais alto de detalhe suportado por um *data warehouse*.

índice Gini métrica usada em economia para avaliar a diversidade da população. O mesmo conceito pode ser usado para determinar a pureza de uma classe específica como resultado de uma decisão para ramificar um atributo/variável especial.

integração de aplicações corporativas (EAI) tecnologia que fornece um meio de transportar dados de sistemas-fonte para um *data warehouse*.

integração de dados integração que abrange três grandes processos: acesso aos dados, federação dos dados e captura de alterações. Quando esses três processos são implementados corretamente, os dados podem ser acessados por e disponibilizados a uma gama de ferramentas de ETL e análise e aos ambientes de *data warehousing*.

integração de informações corporativas (EII) espaço de ferramentas em desenvolvimento que promete integração de dados em tempo real a partir de diversas fontes, como bancos de dados relacionais, *Web services* e bancos de dados multidimensionais.

inteligência de negócios (BI) modelo conceitual para suporte à decisão. Combina arquitetura, bancos de

dados (ou *data warehouse*) ferramentas de análise e aplicações. Ver análise de negócios, *data mining*, processamento analítico on-line (OLAP), sistemas de suporte à decisão (DSS).

Linguagem de Consulta Estruturada (SQL) linguagem de definição e gerenciamento de dados para bancos de dados relacionais. SQL combina com a maioria dos SGBDR.

mapa de características auto-organizável de Kohonen processo abrangente de uso de métodos de *data mining* para encontrar informações e padrões úteis em dados.

mapa estratégico exibição visual que retrata as relações entre os principais objetivos organizacionais para todas as quatro perspectivas BSC.

meta estratégica objetivo quantificado que tem um período de tempo indicado.

metadados dados sobre dados. Em um *data warehouse*, os metadados descrevem o conteúdo de um *data warehouse* e o modo do seu uso.

modelagem dimensional sistema baseado em recuperação que suporta acesso a consultas em grande escala.

momentum parâmetro de aprendizado em redes neurais de retropropagação.

monitoramento de atividades de negócios (BAM) sistema em tempo real que alerta os gerentes sobre possíveis oportunidades, problemas iminentes e ameaças, capacitando-os a reagir com modelos e colaboração.

multidimensionalidade capacidade de organizar, apresentar e analisar dados em várias dimensões, como vendas por região, por produto, por vendedor e por tempo (quatro dimensões).

neurônio célula (ou seja, elemento de processamento) de uma rede neural biológica ou artificial.

núcleo parte central de processamento de um neurônio.

objetivo estratégico declaração ampla ou linha de ação geral que prescreve orientações direcionadas para uma empresa.

OLAP multidimensional (MOLAP) OLAP implementado através de um banco de dados multidimensional (ou *data store*) que resume operações em visões multidimensionais com antecedência.

OLAP relacional (ROLAP) implementação de um banco de dados OLAP sobre um banco de dados relacional existente.

oper mart *data mart* operacional. Um *oper mart* é um *data mart* em pequena escala normalmente usado por um único departamento ou área funcional de uma empresa.

operational data store (ODS) tipo de banco de dados freqüentemente usado como uma área temporária

para um *data warehouse*, especialmente para arquivos com informações de clientes.

perceptron estrutura de rede neural antiga que não usa camada escondida.

peso de conexão peso associado a cada ligação em um modelo de rede neural. Algoritmos de aprendizado para redes neurais avaliam os pesos de conexão.

plano operacional plano que traduz metas e objetivos estratégicos de uma empresa em um conjunto de táticas e iniciativas, exigências de recursos e resultados esperados bem-definidos.

processamento analítico online (OLAP) sistema de informação que permite ao usuário, quando no PC, consultar o sistema, conduzir uma análise e assim por diante. O resultado é gerado em segundos.

processamento paralelo técnica avançada de processamento computacional que possibilita que um computador execute múltiplos processos simultaneamente, em paralelo.

reconhecimento de padrão técnica de combinar um padrão externo com um padrão armazenado na memória de um computador (ou seja, o processo de classificar dados em categorias pré-determinadas). O reconhecimento de padrão é usado em mecanismos de inferência, processamento de imagens, computação neural e reconhecimento de fala.

rede neural artificial (RNA) tecnologia computacional que tenta criar computadores que funcionem como cérebro humano. As máquinas possuem armazenamento simultâneo de memória e trabalham com informações ambíguas. Algumas vezes chamada simplesmente de rede neural. Ver também computação neural.

relatório de status relatório que fornece as informações mais atuais sobre o *status* de um item (p.ex., pedidos, despesas, quantidade de produção).

retropropagação algoritmo de aprendizado mais conhecido na computação neural. Na retropropagação, o aprendizado é feito através da comparação de saídas calculadas às saídas desejadas de casos históricos.

scorecard de desempenho exibição visual que é usada para fazer um gráfico do progresso em relação às metas e aos objetivos estratégicos e táticos.

Six Sigma metodologia de gerenciamento de desempenho com objetivo de reduzir o número de falhas em um processo de negócios para o mais próximo de zero falhas por milhão de oportunidades (DPMO) possível.

sinapse ligação (onde os pesos estão) entre os elementos de processamento em uma rede neural.

sistema de controle de diagnóstico sistema cibernético que tem entradas, um processo para transformação de entradas em saídas, um padrão ou marca comparativa contra a qual comparar as saídas, e um canal de *feedback* para permitir que informações sobre variâncias entre as saídas e o padrão sejam comunicadas e agilizadas.

sistema de informação geográfica (GIS) sistema de informação que usa dados especiais, como mapas digitais. Um GIS é uma combinação de texto, gráficos, ícones e símbolos em mapas.

sistema de medição de desempenho sistema que auxilia os gerentes no rastreamento de implementações da estratégia de negócios através da comparação entre resultados reais e metas e objetivos estratégicos.

sistemas de posicionamento global (GPS) dispositivos sem fio que usam satélites para permitir aos usuários detectar a posição na Terra de itens (p.ex., carros ou pessoas) nos quais os dispositivos estão anexados, com precisão razoável.

sistemas de suporte à decisão modelo conceitual para um processo de suporte à tomada de decisão gerencial, em geral, através da modelagem de problemas e emprego de modelos quantitativos para análise de solução.

suporte automatizado à decisão (ADS) sistema baseado em regra que fornece uma solução para um problema gerencial repetitivo. Também conhecido como gerenciamento de decisões empresariais (EDM).

tabela de dimensão tabela que aborda *como* os dados serão analisados.

taxa de aprendizado parâmetro para aprendizado em redes neurais. Determina a parcela da discrepância existente que deve ser compensada.

teoria da ressonância adaptativa (ART) método de aprendizado não supervisionado criado por Stephen Grossberg. ART é uma arquitetura de rede neural planejada para ser semelhante a um cérebro em modo não controlado.

teste de caixa preta teste que envolve comparação entre resultados de teste e resultados reais.

text mining aplicação de *data mining* a arquivos de texto não estruturados ou menos estruturados. Envolve a geração de índices numéricos expressivos a partir do texto não estruturado e, em seguida, o processamento desses índices usando vários algoritmos de *data mining*.

topologia maneira na qual os neurônios são organizados em uma rede neural.

valor limite valor de barreira para a saída de um neurônio acionar o próximo nível de neurônios. Se um valor de saída for menor do que o valor limite, não será passado para o próximo nível de neurônios.

visão estratégica idéia ou imagem mental de como a empresa deve se parecer no futuro.

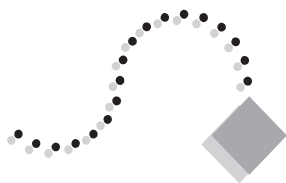
visualização de dados apresentação gráfica, em animação ou vídeo de dados e dos resultados da análise de dados.

Web mining descoberta e análise de informações úteis e interessantes a partir da Web, sobre a Web e, em geral, através de ferramentas baseadas na Web.

Web mining de conteúdo extração de informações úteis a partir de páginas Web.

Web mining de estrutura desenvolvimento de informações úteis a partir de *links* incluídos em documentos da Web.

Web mining de uso extração de informações úteis a partir de dados que são gerados por visitas a páginas Web, operações etc.



Índice

Nota: os números de página seguidos por *t* ou *f* referem-se a tabelas ou figuras. Aqueles seguidos por um A ou T referem-se às caixas de Casos de Aplicação ou *Insights* de Tecnologia, respectivamente.

A

- Abordagem “plano grande, construção pequena”, 78
- Acesso a fontes de dados, 70
- Acesso aos dados, 70
- Acidentes com veículos automotores, 157-159A
- Acurácia preditiva, 162-164
- Administrador do *data warehouse* (ADW), 91
- Advanced Software Applications, 172-173
- ADVISOR *Solutions*, 138-139
- ADVIZOR *Solutions*, 125-126
- Aer Lingus, 178-180A
- Agência de Segurança Nacional (NSA), 174, 176-177A
- Agentes inteligentes, 136, 162-163
- AGIS, 131
- Agrupamento, 156-158*t*
- Agrupamento hierárquico, 170-171
- Albuquerque, cidade de, 230-231A
- Alerta, 106-107
- ALG *Software*, 222*t*
- Algoritmos, 135-136, 155-159, 162-169
- Algoritmos CART/SPRINT, 164-169
- Algoritmos genéticos, 162-163
- Alinhamento estratégico, 205T
- Allied Building Products Corp., 110-111
- Allied Signal, 214
- Amazon.com, 136, 152, 183, 184*f*, 185
- Ambeo, 92T
- Ambiente de negócios
 - complexidade para as empresas, 21, 24-25
 - mudanças e suporte computadorizado à decisão, 23-24
 - principais fatores, 24-25*t*
 - reações organizacionais, 24-26
- Ambientes de rede inteligentes, 193
- Ameaça terrorista, 160-161A, 174, 176-177A
- Análise, 34-35
- Análise da situação atual, 196-198
- Análise da Web, 137-140, 158-159A
- Análise de acesso na Web, 41-42, 137-138, 183, 185
- Análise de cesta de supermercado, 156
- Análise de cliente, 158-159A
- Análise de *cluster*, 162-163, 169-171
 - hierárquica, 170-171
 - métodos e exemplo, 170-171
 - software*, 172-173
- Análise de conteúdo, 182-183A
- Análise de cubo, 106-107, 118
- Análise de exceção, 200-202
- Análise de *hiperlink*, 182-183A
- Análise de lacuna, 196-198
- Análise de *link*, 182-183A
- Análise de negócios (BA), 29-31
 - aplicações avançadas, 105-106, 120-123
 - dados de entrada para, 101
 - definida, 104-105
 - ferramentas e técnicas, 105-109
 - ferramentas listadas, 123
 - fundamentos, 104-106
 - por que projetos falham, 143-144
 - sucesso e utilidade, 141-144
 - usando a Web, 137-138
 - uso, 140-142
- Análise de *site*, relatórios, 115*t*
- Análise do setor, 41-42
- Análise *drill-down*, 80-81, 108-109
- Análise estatística e *data mining*, 106-107
- Análise preditiva, 31-34A, 121-123A
 - data mining* e, 121-123
 - ferramentas, 121-122
 - identificando padrões de comportamento, 122-123
- Analítica, 125-126
- Angoss Software, 123, 138-140*f*, 171-172*t*, 184*t*
- Aplicação analítica, atividades disponíveis, 104-105
- Aplicação de modelagem de crédito, 125-126*f*
- Aplicações de fluxo de trabalho, 135-136
- Applix, 222*t*
- Apresentação multidimensional, 117-118
- Arkansas, Estado de, 146-148A
- Armadilhas da análise de variância, 200-202
- Armazenamento, 37-39
- Arquitetura, 28-32, 30*f*. *Ver também* Arquitetura de *data warehousing*; *Tipos específicos*
 - fatores que afetam a seleção, 68-69
 - modelos com código aberto, 180-181
- Arquitetura cliente/servidor, 58-59
- Arquitetura de *data warehouse* distribuída, 66-67*f*
- Arquitetura de *data warehousing*, 61-70
 - arquiteturas alternativas, 64-70
 - duas e três camadas, 61-63*f*, 64*f*
- Arquitetura de *data warehousing* empresarial, 64-65*f*
- Arquitetura de metanegócios, 59
- Arquitetura de sistema, 217
- Arquitetura em estrela, 67-68
 - RDW comparado, 88
- Arquitetura em estrela de *data mart*, 65-66*f*
- Arquitetura orientada a serviços (SOA), 47
 - dados táticos e estratégicos misturados sob demanda, 90-91
 - metadados, 59
 - serviços de granularidade grossa, 70-71

Arrecadação de impostos, 33-34A
 Árvores de decisão, 155-158A, 161-169
 Ascential, 45-46
 Associações, 156-158t
 AT&T, 86
 Ativo, conceito de, 90-91T

B

BA. *Ver* Análise de negócios
Balanced scorecard (BSC), 210-213
 alinhando estratégias e ações, 210-213
 certificação, 213
 como metodologia, 30-31
 definições e funções, 210-211f
 para medida de desempenho, 206
 Six Sigma comparado, 216T
 Balanço, em BSC, 210-211
 Banco, 87A, 113A, 122-123, 128-129
 Banco de dados abrangente, 61-62
 Banco de dados relacional distribuído, 109-110
 Bancos de dados
 camada de banco de dados do sistema BPM, 217-220
 multidimensional, 117
 sistemas de BPM, 219-220
 Bancos de dados operacionais, 36-37
 Bancos de dados transacionais, 121-122
 estrutura de *data warehousing versus*, 83-84
 Bancos online, 87A
 Bank of America, 70-71A
 BEA, 223-224T
 Ben & Jerry's, 105-106A
 Benefícios concretos
 definidos, 85-86
 Bethune, Gordon, 55-56
 BI. *Ver* Business Intelligence
 Big Horn Computer Services, 131
 Blue Cross e Blue Shield, 150-151
 Blue Rhino, 111t
 BPM Standards Group, 217
 British Airways, 111t
 British Petroleum Exploration and Production, 111t
 Brobst, Stephen, 90-91T
 Bromford Iron & Steel Co. Ltd., 111t
 BSC Collaborative, 206-207, 213
 BSC. *Ver* Balanced scorecard
 Business Objects, 76t, 91, 110-111, 114, 118, 120, 132-133, 137-138, 142-144, 143A, 222t, 227-229A
 Business Intelligence (BI)
 análise de negócios e, 104-105
 aplicações comuns, 31-32
 arquitetura, 28-32, 30f
 benefícios, 47
 BPM comparado, 194, 195t

 como suporte computadorizado à decisão gerencial, 21
 componentes, 28-32
 comunidades de usuários, 42-43t
 controle, 34-36
 definições, 26-27
 desenvolvendo ou adquirindo sistemas de BI, 45-46
 estrutura para, 25-35
 exemplos de aplicações típicas, 28-29t
 história, 26-28f
 implementação, 42-47
 integração entre sistemas e aplicações, 45-47
 justificativa e análise de custo/benefício, 45-46
 obrigação estratégica de, 39-41
 origens e motivações, 27-28
 por que projetos falham, 143-144
 problemas de sistemas antigos, 23-24
 relação custo/benefício, 34-35
 retorno do investimento, 22-23
 segurança e proteção da privacidade, 45-46
 teorias de, 37-39
 valor comercial de, 33-34
 Business Intelligence (BI) em tempo real, 44-46
 Continental Airlines, 55-57
 estritamente relacionada à análise de negócios, 131-133
 Business Intelligence sob demanda, 45-46
 Business performance management (BPM), 26-27
 aplicações, 219-221
 arquitetura, 217-220, 218f
 BI comparada, 194, 195t
 como componente de BI, 30-32
 como próxima fase, 191
 comparações comuns feitas, 228-229
 definido, 30-32, 194
 insuficiência analítica, 204, 206
 interface de usuário, 221-222
 metodologias, 210-217
 balanced scorecards, 210-213, 216T
 Six Sigma, 213-217, 216t
 monitoramento, 200-204
 pacotes comerciais, 222-224
 processos resumidos, 195-196, 195f
 testando melhores práticas, 202-204, 206, 205T

C

Cadeia de patrocinadores, 83-84
 Camada de aplicações, de sistemas de BPM, 217-218

Camada de interface cliente ou de usuário do sistema BPM, 217-218
 Capital One, 152
 Captura de alterações, 70
 Carregamento, 72
 Carregamento de dados, 61-62
 Cartesis, 222t
 Casos (histórias de sucesso do cliente), 92-93
 Celequest, 232-233
 Celerant Consulting, 215
 Centro de competência de BI (BICC), 44
 exemplo France Telecom, 44-45A
 CI. *Ver* Inteligência competitiva
 Ciclo dos negócios
 tempo reduzido, 27-28
 Ciclo OODA, 236-238A
 Cidade de Nova Iorque, 130A
 CIF. *Ver* Customer information file
 CIO Insight, 141-142T
 Cisco Systems, 192-194, 198
 Clarity Systems, 222t
 Classificação, 155-158t, 162-164
 ClickTracks, 184t
 Clustan, 172-173
 CNN, 86
 Coca-Cola Bottling Company (Hokuriku, Japão), 74-76A
 Coda, 222t
 Coffee, Peter, 90-91T
 Cognos Inc., 110-111, 115, 118, 120, 123, 132-133, 137-138, 142-143A, 171-172t, 222t, 230-233, 231A
 Coldwater Creek, 142-143A
 Coletores, 73-76
 Comércio eletrônico, 138-139A
 Companhia de Seguros Fireman's Fund, 179-180
 Comportamento do cliente
 data mining, 158-159A
 CompStat, 130A
 Compudigm International, 124-125
 Computer Associates, 76t, 110-111
 Comunidades de usuários, 42-43t
 benefícios considerados, 73-76
 orientação do gerente de banco de dados, 83-84
 participação no desenvolvimento do *data warehouse*, 85-86
 Conceito de tempo real, evolução do, 90-91T
 Condutor do projeto, 83-84
 Configuração de produto ou serviço, 133-134
 Conformidades corporativa e reguladora, 135-136
 Consonus, 124-125
 Consultas *ad hoc*, 116-117
 e análises, 106-107

Continental Airlines, 54-56, 89-90, 93-94
 Controle operacional, 135-136
 Cooperativa de Crédito do Texas, 128
 CorVu, 222t
 Criação e uso de inteligência, 34-36f
 Crime do colarinho branco, 141-143
 Cross-Industry Standard Process for Data Mining (CRISP-DM), 172-174f
 CSX Transportation Inc., 130A
 Cubo, definido, 118
 Cubos com visões parciais, 119f
 Cubos de dados
 definidos, 118
 multidimensionais, 118, 119f
 Cumprimento da lei, 174, 176-177A, 182-183, 185A
 Custeio baseado em atividade (ABC), 219-221
 Custo total de propriedade (TCO), 34-35, 224T
 Customer information file (CIF), 58-59

D

Dados de acesso na Web, 137-138
 Dados de entrada, 37-39
 Dados de saída, 37-39
 Dados e regras de negócios, 73
 Dados estáticos, 44-45
 Dados não-voláteis, em *data warehousing*, 58-59
 Dados quase em tempo real, 87A
 Dados quase em tempo real de cada ponto de venda (PDV), 74-76A
 DASD. Ver Dispositivos de armazenamento de acesso direto
 Dashboards
 BSNF, 51-53A
 business performance management, 31-32
 como ferramentas para geração de relatório ampliado, 115
 corporativo, 236-237A
 estratégicos, 227-228
 fábrica, 236-237A
 indicadores-chave de desempenho, 31-32
 operacionais, 227-228
 poder do BI em tempo real, 89-90
 recursos Hyperion, 22-23
 relatórios iniciais de indicadores-chave, 113A
 táticos, 227-228
 usos, 47
 versus scorecards, 225-228, 226t
 visualização, 125-126
 Dashboards de desempenho, 222, 225-230
 códigos de cores, 228-229
 dashboard típico, 222, 225f

definidos, 225-228
 o que procurar, 229-230
 para médicos, 227-229A
 principais tipos, 227-228
 projeto, 228-229
 scorecards comparados, 225-228, 226t
 Data Advantage Corp., 76t
 Data marts, 29-30
 abordagem de *data mart*, 77-80, 77t, 78t
 aplicação de BPM, 218-220
 arquitetura, 65-66f
 arquitetura de barramento com *data marts* dimensionais vinculados, 66-67
 consolidando em EDW, 78-80A
 construir sobre, 47
 definidos, 58-59
 EDW relacionado, 60-61, 78-80A, estrutura, 64-65
 estrutura avaliada, 70
 federados, 67-69
 Data marts de aplicações, 218-219
 Data marts dependentes
 definidos, 58-59, 67-68f
 Data marts independentes, 66-68f
 definidos, 58-59
 Data mining
 algoritmos, categorias, 155-159
 análise de *cluster* para, 169-171
 análise preditiva e, 121-123
 aplicações, 159-160
 aumento da popularidade, 152-153
 baseado em descoberta, 158-159
 baseado em hipótese, 158-159
 características e benefícios, 152-155
 como funciona, 155-159
 comportamento do cliente, 158-159A
 conceitos e aplicações, 152-153
 definido, 30-31, 152-153
 empresas de assistência médica, 151-152
 erros crassos, 175T
 ferramentas, 121-122
 inteligência competitiva, 136
 métodos para identificar padrões em dados, 155
 MicroStrategy, 25-26, 106-107
 mitos, 175T
 OLAP comparado, 110-111
 processos para projeto, 172-177
 software, 170-172
 técnicas e ferramentas, 161-173
 usos para, 152-153
 Data Mirror, 132-133
 Data stores transacionais, 218-219
 Data warehouse (DW)
 arquitetura de *data warehouse* distribuída, 66-67f
 ativo, 74-76A

ativo vs. dados históricos, 131-132
 baseado na Web, 85-86
 benefícios, diretos e indiretos, 73-74
 centralizado, 67-68f
 como componente de BI, 28-30
 construindo sobre, 47
 definido, 21-22, 57-58
 em tempo real, 131-133
 escalabilidade, 85-86
 estrutura (esquema estrela), 60-61f, 78-82f
 estrutura e visualizações, 60-61f
 fornecedores, 74-76t
 grão, 80-81
 hospedado, 78-81T
 integrado, 70-71A
 OLPT comparado, 36-37
 principais componentes, 60-62
 progressão de informação em, 73
 questões de administração e segurança, 91-92
 sistemas BPM, 218-220
 suporte à alta direção, 85-86
 suporte à tomada de decisão, 37-39
 tipos de
 data marts, 58-59
 data warehouse empresarial (EDW), 59
 operational *data stores*, 58-59
 Data warehouse de seguro de automóveis, 81-82f
 Data warehouse em estrela, 67-68f
 Data warehouse empresarial (EDW)
 abordagem de EDW (Inmon), 76-80, 77t, 78t
 Continental Airlines, 55-59
 definido, 59
 maneiras de construir, 67-68f
 Data warehouse VISION, 59-61A
 Data warehouses gigantescos, 85-86
 Data warehousing
 baseado na Web, 62-65, 64f
 benefícios, 82-83
 características, 57-59
 como disciplina, 58-59
 como processo contínuo, 84-85
 estratégia para, 74-76
 fases do projeto, 83-84
 fatores de fracasso, 83-84
 First American Corporation, 59-61A
 fontes de dados, 60-62
 fracassos, pesquisa sobre, 82-84
 implementação, 80-86
 Business Intelligence e, 37-39
 melhores práticas, 82-83
 modelo de maturidade, 82-83
 modelo para o sucesso, 84-86
 projeto de banco de dados transacional *versus*, 83-84

- referenciais, 82-83
 - resumo de pré-requisitos, 84-85
 - riscos, 82-84
 - tarefas essenciais que poderiam ser executadas em paralelo, 81-83
 - terceirização no exterior, 86
 - três tipos de produto, 76
 - visão geral do processo, 59-62
 - Data warehousing* ativo (ADW), 74-76A, 87
 - data warehouse* tradicional comparado, 90-91*t*
 - definições, 90-91
 - evolução das decisões, 87-89, 88*f*, 90*f*
 - fábrica de informações corporativa comparada, 90-91
 - realidades em tempo real, 90-91T
 - Data warehousing* baseado na Web, 62-65, 64*f*
 - Webhousing*, 85-86
 - Data warehousing* em tempo real (RDW), 86-91, 131-133
 - comparado a OLPT, 88
 - Continental Airlines, 55-58
 - data warehouse* tradicional comparado, 88-91*t*
 - Overstock.com Inc., 96-98A
 - programas mais recentes, 58-59
 - questões, 89-90
 - sistemas de suporte à decisão com, 57-58
 - Data Warehousing Institute, 74-76, 82-83, 92
 - Database OLAP (DOLAP), 110-111
 - Databeacon Inc., 132-133
 - DataMirror Corp., 76*t*
 - DB2, 64-65, 116-117
 - Decisões de encaminhamento ou segmentação, 135-136
 - Decisões de segmentação, 135-136
 - Declaração de renda reversa, 202-204A
 - Decomposição em valores singulares, 189-190A
 - DeepMetrix, 184*t*
 - Defeitos por milhão de oportunidade (DPMO), 216-217
 - Deluxe Corp., 111*t*
 - Departamento de Comércio dos EUA, 160-161A
 - Departamento de Defesa dos EUA, 85-86
 - Departamento Nacional de Reconhecimento (NRO), 174, 176-177A
 - DER. *Ver* Diagramas entidade-relacionamento
 - Descoberta de conhecimento em bancos de dados (KDD), 173-174, 176-177
 - Descoberta de informações e conhecimento, 105-107*f*
 - Descoberta de sequência, 156-158*t*
 - Desenvolvimento de *data warehouse*, 73-86
 - abordagens, 76-80
 - categorias de benefícios, 73-76
 - considerações adicionais, 78-80
 - fatores perceptíveis e menos perceptíveis, 92
 - Desktop OLAP, 110-111
 - Deteção de fraude, 135-136, 141-143
 - Deteção de fraude financeira, 161-162
 - Determinação de preços inteligente, 30-31A
 - Diagramas entidade-relacionamento (DER), 77
 - Diferenciação, 41-42
 - DigiMine, 142-143A
 - Dimensões, definidas, 117
 - Diretor financeiro (CFO), 30-31, 125-126
 - Disneyland Paris, 201-204A
 - Dispositivos de armazenamento de acesso direto (DASD), 85-86
 - DM Review*, 74-76, 92-93
 - Dow Elanco, 129
 - DSS. *Ver* Sistemas de suporte à decisão (DSS)
 - Duke University Hospital, 215
 - Dun & Bradstreet, 111*t*
 - DuPont, 214
- E**
- EAI. *Ver* Integração de aplicações corporativas
 - E-business*, 87
 - Edison, Thomas A., 174, 176-177
 - Egg Plc, 87A
 - EIS. *Ver* Sistemas de informações executivas
 - Elity, 232-233
 - Embarcadero Technologies, 73-74, 76*t*, 92T
 - Emergency Medical Associates (EMA), 227-229A
 - Emergency Medicine and Analysis Report System (EMARS), 227-229A
 - Empresa de latência zero, 229-230
 - Empresas
 - avaliação, 205T
 - como elas “cultivam” a utilização de dados, 24-26
 - respostas aos ambientes de negócios, 24-26
 - tomada de decisão estratégica, 21
 - Empresas de assistência médica, 150-152
 - Empresas de modelagem de informações, 65-66
 - Endeca, 125-126
 - Entrega de relatórios e alertas, 106-107
 - recursos de *software*, 116-117
 - Entropia, 168
 - Era “.com”, 192, 196
 - Escalabilidade de *data warehouses*, 85-86
 - Especificações de operação proforma, 202-204A
 - E-spionagem, 136
 - Espionagem, 35-37, 137
 - Espionagem industrial, 137
 - Esquema estrela, 78-82*f*
 - ESRI, 131
 - Estoque, 101-103
 - Estratégia, negócios, 196-198
 - definições, 196
 - planejamento e alinhamento com, 42-44
 - Estrutura de *data warehouse* para toda a empresa, 64-65
 - Estrutura do cubo, 110-111
 - Estrutura multidimensional, em *data warehouses*, 58-59
 - Estrutura relacional, em *data warehouses*, 58-59
 - Estudos de caso
 - acidentes com veículos automotores e distrações do motorista, 157-159A
 - Aer Lingus, 178-180A
 - Albuquerque, cidade de, 230-231A
 - análise preditiva, 33-34A, 121-123A
 - Arkansas, Estado de, 146-148A
 - Bank of America, 70-71A
 - Ben & Jerry's, 105-106A
 - Cisco Systems, 192-194
 - comércio eletrônico, 138-139A
 - comportamento do cliente com *data mining*, 158-159A
 - Continental Airlines, 54-56
 - data mining* para cumprimento da lei, 174, 176-177A
 - data warehouse* integrado, 70-71A
 - determinação de preços inteligente, 30-31A
 - Egg Plc, 87A
 - Emergency Medicine and Analysis Report System (EMARS), 227-229A
 - Euro Disney, 201-204A
 - Ferrovias Burlington Northern e Santa Fe Pacific (BSNF), 51-53A
 - First American Corporation, 59-61A
 - First Health Group Corp. (FHG Corp.), 153-154A
 - France Telecom, 44-45A
 - grupos extremistas na Internet, 182-183, 185A
 - Hewlett-Packard Company, 78-80A

- Hokuriku Coca-Cola Bottling Company, 74-76A
 informação para o serviço secreto, 160-161A
 International Truck and Engine Corporation, 207-209A
 J.Crew, 158-159A
 Lexmark International, 101-103
 Longs Drug Stores, 30-31A
 Merrill Lynch, 127A
 Newport News, 138-139A
 Overstock.com Inc., 96-98A
 portal de KPI (baseado na Web), 207-209A
 predição de engarrafamento, 121-123A
 Rastreamento por SIG e GPS, 130A
 Seguro-saúde, 150-152
 TCF Financial Corp., 113A
 Texas, estado do, 33-34A
text mining na Hewlett-Packard Company, 188-190A
 Toyota Motor Sales (EUA), 21-23
 tratamentos médicos, 159-161A
 uso de BI no varejo, 142-143A
 Western Digital (WD), 236-238A
 Euro Disney, 201-204A
 Eventos de desempenho, 230-231A
 Eventos de notificação, 230-231A
 Eventos operacionais, 230-231A
eWeek, 90-91T
 Executivos de TI, uso de ferramentas de BI e BA, 141-142T
 Exigências do consumidor
 fatores do ambiente de negócios, 24-25t
 Extensible Markup Language (XML), 67-71, 213
 Extração, 72
 Extração de dados, 61-62
 Extração de informação, 178-181
 Extração de termos, 179-181
- F**
 Fábrica de informação
 conceito de, 37-39
 exemplo, 38-39f
 Fábrica de informações corporativa, 90-91
 Fábrica de informações da empresa, 37-39
 Fair Isaac, 123
 Fator crítico de sucesso (FCS), 108-109t
 no planejamento estratégico, 196-198
 Fatores sociais
 ambientes de negócios relacionados, 24-25t
 FCS. *Ver* Fator crítico de sucesso
 Fechamento virtual, 192-194
- Federação de dados, 70
 Ferramentas de gestão de dados transacionais, 96-97A
 Ferramentas de planejamento de recursos empresariais (ERP), 106-109
 Ferramentas de transformação, 73
 Ferramentas de transformação de dados, 73
 Ferramentas para reconhecimento de padrões, 160-161A
 Ferrovias Burlington Northern e Santa Fe Pacific (BSNF), 51-53A
 First American Corporation, e *data warehousing*, 59-61A
 First Health Group Corp. (FHG Corp.), 153-154A
 FirstRain, 232-233
 Ford Motor Co., 111t, 214
 ForestTree 6.0, 125-126
 Fornecedores, *data warehouse*, 74-76
 instruções para escolha, 76
 listados, 76t
 Fornecedores da tecnologia de núcleos, 65-66
 France Telecom, 44-45A
 Frequência do termo (fator *tf*), 179-181
 Frequência inversa do documento (fator *idf*), 169
- G**
 Ganho de informação, 168
 Garcella, Jack, 97-98A
 Gartner Inc., 194, 219-220, 222, 229-230
 relatório, 42-44
 Geas, 110-111
 General Architecture for Text Engineering (GATE), 180-181
 General Electric, 214, 215
 Geostor, 146-148A
 Gerenciamento
 e estratégia de longo termo, 198
 por exceção, 200-201
 tempo gasto em estratégia, 204, 206
 Gerenciamento de desempenho corporativo (CPM), 143-144, 194. *Ver também* Sistemas de gerenciamento de desempenho
 Gerenciamento de lucro, 133-134
 Gerenciamento de relacionamento com o cliente, 36-37
 BI em tempo real, 44-45
 data mining, 153-154A
 relatórios, 115t
 Gerenciamento de rendimento, 133-134
 Gerenciamento de risco, 125-126f
 Gerenciamento por exceção, 200-201
- Gestão da cadeia de suprimento (SCM), 36-37
 BI em tempo real, 44-45
 relatórios, 115t
 Gestão de decisão empresarial (EDM), 132-133. *Ver também* Suporte automatizado à decisão
 Gestão de empresas estratégica (SEM), 194
 GIS e GPS, 127-131
 uso por governo de estado, 146-148A
 GlaxoSmithKline, 111t
 Globalização, 39-41
 GoldenGate Software Inc., 96-97A
 Google, 138-139
 Governo da cidade de Richmond (British Columbia), 131-132
 GPS, 130
 Grão (granularidade), 80-81
 Grupo de contrabando, 177A
 Grupos extremistas na Internet, 182-183, 185A
- H**
 Hackett Group, 202-204, 205T, 206-207
 Harrah's Entertainment, 41-42, 124-125
 Harry Rosen Inc., 142-143A
 Harte-Hanks, 159-161A
 Hewlett-Packard Company, 76t, 78-80A, 124-125, 131-132
 text mining, 188-190A
 Highmark Inc., 150-152
 Hilton, 63-65
 Hitachi, 137
 Hokuriku Coca-Cola Bottling Company, 74-76A
 Horizonte de planejamento, 196-198
 Hotscripts, 137
 HTML dinâmica (DHTML), 223-224T
 Hudson's Bay Co., 142-143A
 Hummingbird Ltd., 76t, 137-138
 Hurd, Mark, 78-80A
 HyperActive Technologies, 122-123
 Hyperion Solutions Corp., 22-23, 51-52A, 76t, 110-111, 115, 118, 120, 123, 125-126, 137, 208-209A, 222, 222t
 Hyperion System, 26-27, 223-224T
- I**
 IBM, 51-52A, 73-74, 76t, 110-111, 116-117, 120, 123, 131-133, 137-138, 143-144, 159-161A, 171-172t, 180-181
 Identificação por radiofrequência (RFID), 36-37, 44-45, 136

ILOG, 123

Immon, Bill, 76-80, 77*t*, 78*t*

IMS, 132-133

Indicador, 122-123

Indicadores-chave de desempenho

(KPI), 31-32, 108-109*t*, 113A

objetivos estratégicos, 204, 206

Índice Gini, 164-169

Indução de regra, 156, 162-163

Informação para o serviço secreto, 160-161A

Informatica Corp., 45-46, 73-74, 76*t*,

110-111, 113A, 132-133, 137-138

Information Builders, 110-111, 128

Informix, 109-110, 208-209A

Infra-estrutura, 29-30

fornecedores, 65-66

Infra-estrutura de TI, suporte à integração entre BSC e Six Sigma, 216T

ING Antai Insurance Co. (Taiwan), 110-111

Inrix, 121-123A

Insightful Corp., 123, 180-181

Insights de Tecnologia

data mining

erros crassos, 175T

mitos, 175T

Integração

de *data warehousing*, 57-58

desenvolvimento de sistemas e necessidade de, 143-144

entre sistemas de BI e aplicações, 45-47

Integração de aplicações corporativas

(EAI), 70-71

Integração de dados, 70-72

Integração de informações corporativas (EII), 70-72

metadados, 59

Integração entre médios, 34-36

Inteligência, 35-37

Inteligência artificial, 26-27

possível fusão de BI com, 47

Inteligência artificial de negócios, 28-29

Inteligência competitiva (CI), 39-43, 135-137

Interface gráfica de usuário (GUI), 62-63

Interfaces de usuário, 31-32

Intergraph, 131

International Data Corporation (IDC), 222

International Truck and Engine Corporation, 207-209A

InterNetivity, 118, 120

Intersolve, 110-111

Iteration Software, 45-46

J

J.Crew Group, 142-143A, 158-159A

J.P. Morgan Chase (Nova Iorque), 136

Japão, 74-76A

Java, 223-224T

Just-in-time (JIT)

K

KeySpan, 120-122

Kimball, Ralph, 77-80, 77*t*, 78*t*

KnowHow, 132-133

Knowledge eXtraction ENgines, 123

KPI. *Ver* Indicadores-chave de desempenho

L

Lacuna estratégica, 25-26, 198

Latência, 229-230

Lavagem de dinheiro, 160-161A

Lawson Software, 222*t*

Legislação e regulamentação, 27-28

questões de segurança e privacidade, 91

software para conformidades corporativa e regulatória, 135-136

Lei Sarbanes-Oxley de 2002, 27-28, 33-35, 135-136

Lexmark International, 101-103

Líderes de baixo custo, 41-42

Limpeza dos dados, 84-85

Linguagem de Consulta Estruturada (SQL), 101-102, 116-117

Linksys, 193

Lista de verificação de suposições, 202-204A

Longs Drug Stores, 30-31A

Longview Solutions, 222*t*

Louisiana, Estado de, 128

M

Mapa estratégico, 212-213, 212*f*, 213*t*

Mapas digitais, 122-123A

MapInfo, 131

Máquinas de venda, 74-76A

Marriott International, 152

Matriz de semelhança (distância), 170-171

Mayo Clinic, 159-161A

McDonald's, 231-232

Mecanismos de regras, 231-232

Medidas em multidimensionalidade, 117

Medidas financeiras, desvantagens, 206-207

Megaputer Intelligence Inc., 171-173, 172*t*, 179-180A, 184*t*

Mensagens instantâneas (IM), 44-45

Mentisys, 232-233

Mercado de telecomunicações, 44-45A

Mercados

fatores do ambiente de negócios, 24-25*t*

Merril Lynch, 127A, 214

Meta estratégica, 196-198

Metadados, 59

como característica de *data warehouse*, 58-59, 61-62

Metodologia de análise avançada da

Teradata, 37-40*f*

Metodologia para melhora do processo, 213

Métodos de realidade virtual (RV), 127

Métodos estatísticos para *data mining*, 161-162

Microsoft Corp., 33-34, 73-74, 76*t*, 110-111, 115-118, 121-122, 124-125, 131-

132, 142-144, 143A 189-190A

MicroStrategy, 101-102, 105-107, 110-111, 114, 116-117, 120-123, 143-144,

184*t*

arquitetura MicroStrategy 8, 107-108*f*

Middleware

ferramentas, 61-62

para análise de negócios, 29-30

Migração de dados, 73

ferramentas, 64-65

Modelagem de acesso, 85-86

Modelagem de dados, 85-86

Modelagem dimensional, 77, 78

definida, 78-80

Modelagem e otimização da lucratividade, 219-221

Modelagem preditiva

área da saúde, 152

Modelo de desempenho DMAIC, 172-174*f*, 215

Modelo de indução supervisionada, 155-158*t*

Modelo de pressões de negócios – reações – suporte, 23-26, 24*f*

Modelo definir, avaliar, analisar, melhorar e controlar (DMAIC), 172-174*f*, 215

Modelo DMADV, 215

Modelos MSS, 125-126*f*

Modo “e se”, 51-52A

Monitoramento de atividade de negócios em tempo real, 230-232, 231A

Monitoramento de atividades de negócios (BAM), 229-233

benefícios, 231-232

BI em tempo real, 45-46

fornecedores, 232-233

questões, 231-232

Motorola Inc., 73
 Multidimensionalidade, 117-120
 ferramentas e fornecedores, 118, 120
 limitações, 120
 principal vantagem, 117

N

National Basketball Association (NBA), 160-162
 National Highway Traffic Safety Administration (NHTSA), 157-159A
 Navistar International Corporation, 207-208A
 Newport News, 138-139A
 Nichos de mercado, 41-42
 Nordstrom Inc., 142-143A

O

Objetivo estratégico, 196-198
 Obliquidade, princípio de, 206-208
 Obliquidade da medida, 206-208
 ODS. *Ver Operational data store*
 OLAP. *Ver* Processamento analítico *online*; Sistemas de processamento analítico *online*
 OLPT. *Ver* Processamento de transações *online*
Oper marts, 59
Operational data store (ODS), 58-59
 Oracle, 22-23, 33-34, 64-65, 73-74, 76*t*, 87A, 105-106A, 110-111, 116-117, 131-132, 137-138, 143-144, 146-147A, 222*t*, 227-228A, 230-231*a*
 estrutura de BI, 38-40*f*
 Orçamento, e planejamento financeiro, 199-200
 aplicações de BPM, 219-220
 Orientação por assunto, em *data warehousing*, 57-58
 Otimização, 158-159A
 Otimização do lucro (preço), 133-134
 OutlookSoft, 222*t*
 Overstock.com Inc., 87, 96-98A

P

Pacote de *software* completo, 45-47
 Pacotes específicos para o setor, 135-136
 Padrão, 59
 Parceria, 205T
 Paris Technologies, 113
 Periódicos recomendados, 92-93
 Personalização automática de conteúdo, 116-117
 Pessoas e empresas
 avaliação, 205T
 Pizza Hut, 104-105

Planejamento
 de marco, 202-204A
 dirigido por descoberta, 201-204A
 Planejamento centrado em orçamento, 199-200
 Planejamento centrado em táticas, 199-200
 Planejamento e orçamento financeiro, 199-200
 desempenho monitorado, 200-204
 Planejamento estratégico, 196-198
 Plano de negócios Go Foward, Continental Airlines, 54-56
 Plano operacional, 199-200
 monitoramento de desempenho, 200-204
 Planos de incentivo, 198
 Poindexter, John, 174, 176-177A
 Ponto de divisão, 163-164
 Pontuação de crédito, 104-105
 Portal dos KPI (baseado na Web), 207-209A
 Práticas de planejamento e relatório, 202-206
 eficiência e efetividade, 205T
 insuficiência analítica, 204, 206
 marca comparativa, 202-204, 205T, 206
 Precificação inteligente, 133-134
 Predição de engarrafamento, 121-123A
 Preparação de relatório, 84-85
 análise dos resultados do relatório, 117
 auto-assinatura e distribuição baseada em administrador, 116-117
 consultas *ad hoc*, 116-117
 distribuição de relatório, 116-117
 entrega de relatório e alertas, 106-107
 exemplo de produtos dos fornecedores, 114-115
 Linguagem de Consulta Estruturada (SQL), 116-117
 personalização automática de conteúdo, 116-117
 relatório de termos frequentes, 179-180A
 relatório empresarial, 105-106
 relatório financeiro, 115*t*
 relatório interativo, 115
 relatório OLAP, 114-117
 relatórios *ad hoc* ou sob demanda, 114
 relatórios de rotina, 114
 relatórios em tempo real, 91
 serviços bancários, 113A
 tipos representativos, 115*t*
 Previsão, 135-136, 156-158*t*
 arquitetura de BPM, 219-220
 Previsão dinâmica, 135-136

Previsões contínuas, 204, 206
 Privacidade da informação, 91-92
 Procedimentos de segurança, 91-92
 Processamento analítico, 104-105
 Processamento da análise, 36-40
 Processamento de dados de entrada, 37-39
 Processamento de linguagem natural, 178-179
 Processamento de transações, 36-40
 sistemas, 36-37
 Processamento de transações *online* (OLPT)
 OLAP comparado, 109-111
 Processamento direto, 229-230
 Processamento paralelo, 64-65
 Processo, avaliação, 205T
 Processo de negócios, 216T
 Processo Selecionar, Explorar, Modificar, Modelar e Avaliar (SEMMA), 173-174*f*
 Produção e venda de sorvete, 105-106A
 Produtos de *cockpit*, 125-126
 Protetores, 73-76
 Provedor de serviços de *Business Intelligence* (BISP), 80-81T

Q

Qlik Tech, 113
 Quadrante mágico, 222
 Quantive, 45-46, 232-233
 Questões de gerenciamento de mudanças, 231-232
 Questões de integração, 70-72

R

Raciocínio baseado em casos, 161-162
 RDW. *Ver Data warehousing* em tempo real
 Reconhecimento óptico de caracteres, 136
 Recuperação de informação, 178-179
 Recursos de DSS, 92-93
 Recursos e *links*, 92-93
 Recursos humanos, relatórios, 115*t*
 Redes de Kohonen, 157-158A
 Redes neurais, 30-31, 155, 157-158A, 161-163
 Redundância de dados, 73
 Regras de ação, 133-134
 Regras de análise, 133-134
 Regras de contexto, 133-134
 Regras de exceção, 133-134
 Regras de negócios, 73, 132-134
 mecanismos de regras, 135-136
 Regras se-então-senão, 155-156
 Regressão, 156-158*t*
 Relatório de exceção, 108-109

- Relatório de termos frequentes, 179-180A
- Relatório financeiro, 115*t*
Hyperion System 26-27, 224T
- Relatórios de operação e da gestão da cadeia de fornecimento, 115*t*
- Relatórios de *status*, 103
- Relatórios empresariais, 105-106
- Relatórios financeiros
dirigidos por calendário, 206-207
empecilhos dos dados, 206-207
- Relatórios sob demanda ou *ad hoc*, 114
- Rendimentos de ativos (ROA), 196-198
- Reno-Depot, 142-143A
- Resumo de informação, 178-179
- Retail BI System, 101-102
- Retorno do investimento (ROI)
análise de negócios, 141-142
controle de BI, 34-35
data warehousing, 73-76
marketing online, 138-139A
Toyota, 22-23
- RFID. *Ver* Identificação por radiofrequência
- Rivalwatch.com, 136
- Robôs, ajuda ao cliente, 22-23
- ROLAP. *Ver* Sistemas de processamento analítico *online* relacional
- S**
- SAP, 194
- SAP AG, 106-107-108-109
- SAS Institute Inc., 33-34, 70-71, 73-74, 76*t*, 87A, 110-111, 132-133, 142-143, 161-162, 180-181, 184*t*, 188-190A, 222*t*
- Savvion, 45-46, 232-233
- Scientio, 184*t*
- SCM. *Ver* Gestão da cadeia de suprimento
- Scorecards, 115, 125-126, 206-207
arquitetura, 221
- Scorecards de desempenho, 225-228, 226*t*
- Seagate Technology, 118
- Sears, 142-143A
- SeeRun, 232-233
- Seguro-saúde, 150-152, 204, 206A
- Serviços de análise, 189-190A
- Serviços logísticos da Toyota (TLS), 21-26
ações de BI listadas, 24-26
características do sistema, 26-27
- Servidores de mensagem, 231-232
- Setor aéreo e *text mining*, 178-180A
- Setor de discos rígidos, 236-238A
- Setor de jogos de azar, 124-125
- Shells, 45-46
- Siemens, 76*t*
- Sistema de Informação Vigilante (VIS), 236-238A
- Sistema de previsão e planejamento On-Track, 51-53A
- Sistema Revenue Single Source (BNSF), 51-52A
- Sistemas automatizados de decisão (ADS), 29-30
determinação de preços inteligente, 30-31A
- Sistemas de alerta, 103
- Sistemas de automatização de decisão, 29-30. *Ver também* Sistemas automatizados de decisão
- Sistemas de controle de diagnóstico, 200-201*f*
- Sistemas de gerenciamento de banco de dados relacional (SGBDR), 64-65
- Sistemas de gerenciamento de desempenho financeiro, 33-35
- Sistemas de gestão de desempenho (PMS), 143-144
- Sistemas de informações executivas (EIS), 26-27
definidos, 108-109
recursos, 108-109*t*
- Sistemas de medida de desempenho, 206-209
efetiva, 207-209
problemas com sistemas existentes, 206-208
- Sistemas de processamento analítico *online* (OLAP), 36-37, 109-113
consulta típica, 109-111
data mining comparado, 121-122
database OLAP, 110-111
definidos, 109-110
desktop OLAP, 110-111
12 regras do Codd para OLAP, 112*t*
exemplos de aplicações bem-sucedidas, 110-111*t*
ferramentas, 110-112
características de, 112
representativas, 112
fornecedores, 110-113
método direto de suporte à decisão, 109-110
multidimensional, 110-111
OLPT comparado, 109-111
principais tipos listados, 110-111
regras de avaliação de produto, 112*t*
relacional, 110-111
software de análise de negócios, 29-30
Web OLAP, 110-111
- Sistemas de processamento analítico *online* multidimensional (MOLAP), 110-111
- Sistemas de processamento analítico *online* relacional (ROLAP), 110-111
- Sistemas de suporte à decisão (DSS)
com *data warehousing* em tempo real, 57-58
uso de *data warehouse* empresarial, 59
- Sistemas de suporte executivo (SSE), 108-109*t*
- Sistemas empresariais, 135-136, 231-232
- Sistemas inteligentes, 28-29
- Six Sigma, 213-217
balanced scoreboard comparado, 216T
como metodologia, 172-174*f*
definido, 214-215
limitações, 215, 217
- Six Sigma Business Scorecard, 215, 217
- Smartmoney.com, 127A
- SOA. *Ver* Arquitetura orientada a serviços
- Sobrecarga de medida, 206-207
- Society of Competitive Intelligence Professionals, 41-42, 137
- Software* cliente (*front-end*), 61-62
- Software* de aquisição de dados (retaguarda), 61-62
- Software* de ETL, 61-62
- Software* especializado para BAM, 232-233
- Soluções de auditoria do acesso aos dados, 92T
- Sonic Software, 132-133
- SPSS, 110-111, 123, 127, 142-143, 157-158A, 171-173, 172*t*, 180-181, 184*t*
- SQL Server, 64-65
- SQL. *Ver* Linguagem de Consulta Estruturada
- StatSoft Inc., 123, 171-172*t*, 180-181
- Subaru of America, 111*t*
- Sun Microsystems, 87A, 111*t*
- Sunopsis Data Conductor, 97-98A
- Suporte à decisão e sistemas inteligentes, 105-107*f*
- Suporte automatizado à decisão (ADS), 132-136
- Sybase Inc., 76*t*
- Synthean, 232-233
- Systems Union, 222*t*
- T**
- Tabelas de dimensão, 78-81
- Tabelas temporárias, 73
- Tableau Software, 125-126
- TCF Bank, 110-111, 113A
- TCF Financial Corp., 113A
- Técnicas de AJAX, 224T

Tecnologia

- avaliação, 205T
- fatores do ambiente de negócios, 24-25t

Tecnologia de ETL, 230-231**Tecnologia de modelagem de processo, 230-231****Tecnologias de extração/transformação/****carga (ETL), 72-74, 72f**

- BI em tempo real, 45-46

- metadados, 59

- para camada de banco de dados de BPM, 218-219

Tempo, em multidimensionalidade, 117**Tempo de trânsito, 22-23****Temtec, 110-111, 113, 127, 189-190A****Teradata, 51-52A, 64-65, 67-68, 70-****71A, 74-76A, 76t, 87-91T, 88f, 90f,****96-98A, 142-143A, 171-172t****Teradata Magazine, 38-39****Teradata University Network (TUN)**

- casos disponíveis, 93-94

- portal descrito, 48-49

Terceirização no exterior, 86**Texas, Estado do**

- arrecadação de impostos, 33-34A

Text mining, 177-181

- processo, 179-181

- uso pela HP de, 188-190A

Tibco, 45-46, 73-74, 232-233**Time Warner, 111t****TLS. Ver Serviços logísticos da Toyota****Tomada de decisão**

- análise de negócios, 101

- analítica, 152

- comportamento politicamente ino-
- cente, 83-84

- exemplo, 25-26

- ferramentas em tempo real, 86-87

- suporte computadorizado à decisão,
- 23-26

Toshiba, 214**Toyota Motor Sales (EUA), 21-23****Transação, definida, 37-39****Transformação, 72****Tratamentos médicos, 159-161A****TruServ Corp., 142-143A****TUN. Ver Teradata University Network****U****Ultraex, 130A****Unica, 123****Union Pacific Railroad, 130A****Universidade de Geórgia, 130A****Universidade do Arizona, 182-183, 185A****Unstructured Information Management****Architecture (UIMA), 180-181****UPS, 89-91****V****V(isual Insights), 127****Valor presente líquido, 74-76****Vanguard Group, 63-64****Varejistas, progresso da BI, 142-143A****Variações no tempo (qualidade tempo-****ral) em data warehousing, 58-59****Variâncias operacionais, 200-202f****Variedade, 216-217****Varredura de ambiente, 196-198****Velocidade de carregamento da página,****64-65****Vendas de produto, relatórios, 115t****Ventana Research, 230-231A****Visa Plus, 131****Visão estratégica, 196-198****Visões, 118, 119f****Visual FoxPro, 227-228A****Visualização, 105-107f****com data mining, 156-158t****de dados, 124-127, 162-163****ferramentas, 31-32, 47****baseadas na Web, 124-125****fornecedores e produtos, 127****novos rumos, 125-126****planilhas, 124-125****Visualização de dados financeiros, 125-****127A****Vitria, 45-46****Volume de vendas, 101-103****relatório, 115t****W****Wal-Mart Stores Inc., 85-86, 142-143A,****196-198****Walt Disney Company, 201-204A****Warehouse****definido, 37-39****Warehouse empresarial e data store ope-****rationa, 66-67f****Web mining, 181-184, 181f****Web mining de conteúdo, 181-183, 181f****Web mining de estrutura, 181f, 182-183****Web mining de uso, 181f, 183, 184t, 185****Web OLAP (WOLAP), 110-111****webMethods, 45-46****WebTAS, 177A****WebTrends, 138-139, 184t****WEKA, 171-172****Western Digital (WD), 236-238A****Williams-Sonoma Corp., 142-143****World Bank, 111t****World Wide Web/Internet****análise de negócios na, 137-138****BI em tempo real, 44-46****data warehouses, 58-59****data warehousing baseado na Web,****62-65, 64f****impacto competitivo, 39-41****inteligência competitiva, 136****relatórios para os clientes, 115****reunir inteligência, 36-37****SIG, 131****sistemas de BPM, 221****URLs dos websites, 22-23n****visualização de dados, 124-125****website para este livro, 48-49****X****XML (Extensible Markup Language),****67-71, 213**