

As Ferramentas de Business Intelligence

O B J E C T I V O S

- Conhecer as várias ferramentas tecnológicas de *business intelligence* e as suas características básicas
- Diferenciar e segmentar as tecnologias de acordo com os seus objectivos de utilização e importância para determinadas tarefas

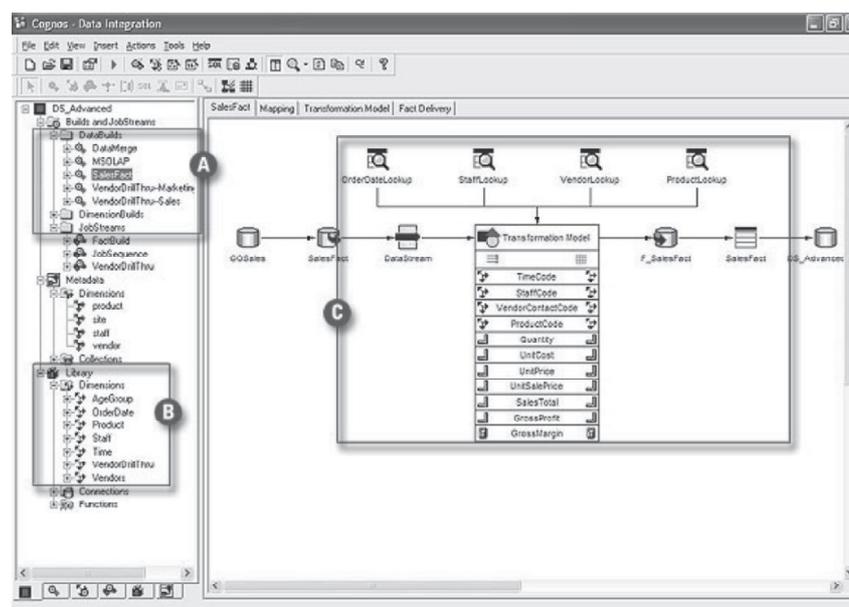
As ferramentas ou aplicações de business intelligence são muitas e variadas. As suas diversas classificações não têm fronteiras estanques. Iremos, contudo, como critério, descrever as mais relevantes e de utilização mais generalizada.

DATA INTEGRATION/ETL

É cada vez mais importante, para o processo de ETL, a utilização de ferramentas de auxílio. Embora tenham sofrido uma enorme evolução, estas ferramentas possuem ainda algumas limitações, pelo que, dependendo da complexidade das transformações, por vezes é necessário parametrizá-las, utilizando linguagem de programação, para completar algumas lacunas que ainda possam existir. No entanto, prescindir destas ferramentas significa uma maior manutenção do processo, à medida que o *data warehouse* vai evoluindo e se vai adequando aos novos desafios do negócio.

As novas tendências nas tecnologias de ETL vão no sentido do desenvolvimento de ferramentas mais poderosas, que, para além das capacidades tradicionais, se apresentem como mais-valias em termos de perfil de dados (*data profiling*), qualidade de dados (*data quality*) e gestão de metadados.

Figura 4.1
Imagem de ecrã
da aplicação
de ETL Cognos
Decision Stream®



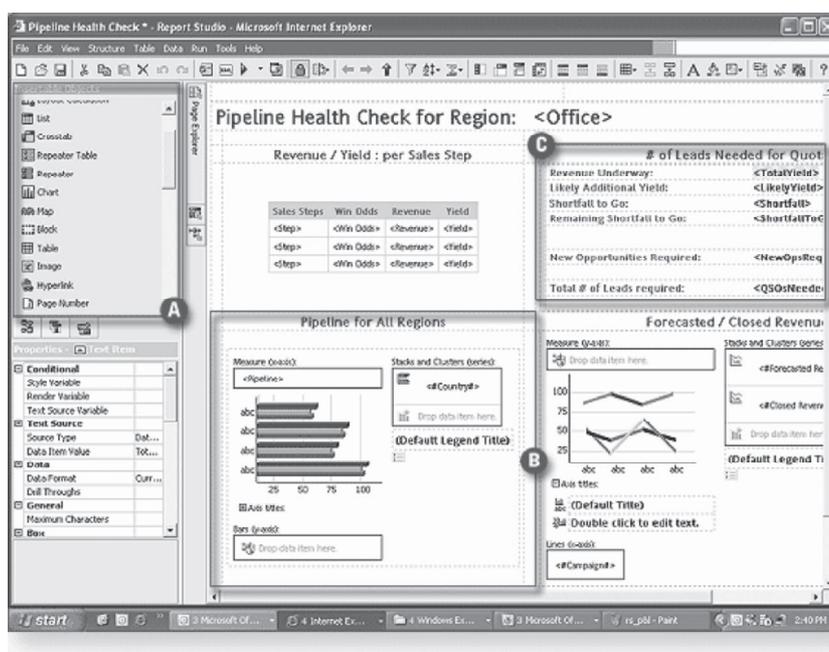
BUSINESS/PRODUCTION REPORTING

As aplicações de *production reporting* justificam-se pela monitorização e pelo acompanhamento das operações da organização numa base de continuidade – muitas vezes diária. As aplicações de BI terão, neste contexto, de assimilar e transmitir, com grande regularidade, os dados fundamentais (quantidades, valor, alertas para valores de excepção) ao acompanhamento e análise.

O principal factor de sucesso destas aplicações será o seu *modus operandi* com as aplicações de *frontoffice* de registo das actividades.

As actividades de produção intensiva, como sejam uma indústria transformadora, uma cadeia de lojas ou uma actividade extractiva, são as que terão maior necessidade deste *reporting* constante.

Figura 4.2
Imagem de ecrã
da aplicação de
Cognos Report
Studio®



FERRAMENTAS DE ANALISYS & QUERYING

Este tipo de aplicações está concebido com o intuito de proporcionar aos utilizadores finais-decisores e analistas altos níveis de autonomia e interactividade no acesso à informação. Os destinatários da informação não ficam pois constrangidos apenas a parcelas segmentadas e/ou parametrizadas do conhecimento útil (*reportings* ou cubos

de informação uniformizados). Poderão, com as funcionalidades destas aplicações, «colocar novas questões» e solicitar novas análises, baseadas em diferentes critérios, e obter assim um maior valor acrescentado das fontes de dados.

Figura 4.3
Imagem de ecrã
da aplicação
Cognos Query
Studio®

The screenshot shows the Cognos Query Studio interface. The main report area displays a table titled 'Product Line Report' for 'Digital2Go'. The table is titled 'Volume and Costs' and has the following data:

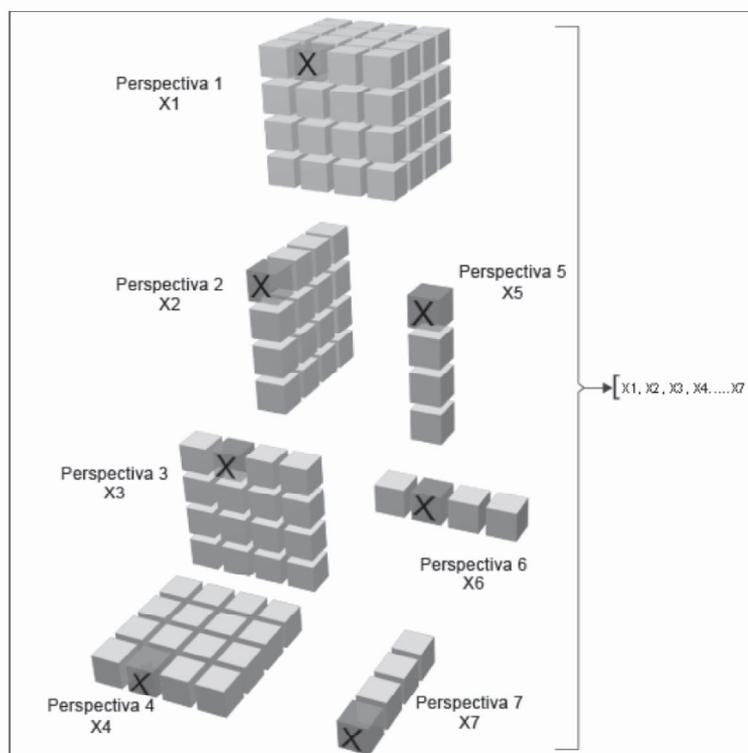
Product Line	Product Type	Expected Volume	Projected Costs
Consumer Electronics	Computer Accessories	28,203	\$434,612
	Computers	7,955	\$441,934
	MP3	13,305	\$1,559,770
	PDA	7,543	\$1,874,760
Consumer Electronics		90,337	\$60,355,256
Entertainment Media	CD Audio	54,466	\$17,983,448
	DVD Video	53,546	\$398,020
	Entertainment Accessories	193,605	\$4,828,291
Entertainment Media	Game Console	72,302	\$14,653,640
	Software	111,896	\$14,318,904
Entertainment Media		485,815	\$52,182,303
Home Office	Chairs	214,871	\$84,298,405
	Desks	170,094	\$81,713,030
	Office Accessories	59,424	\$61,106
Home Office		444,389	\$166,072,541
Home Theatre	Big Screen TV	17,812	\$86,643,000
	Speakers	61,603	\$46,619,730
	Standard TV	9,522	\$12,492,800

APLICAÇÕES ANALÍTICAS (ANÁLISE OLAP)

OLAP (online analytical processing) é um conceito que se refere a aplicações informáticas que permitem efectuar, de forma rápida e partilhada, a análise de informação multidimensional, originária de diversas fontes de dados.

As tecnologias OLAP permitem reestruturar os dados de uma base de dados relacional numa perspectiva multidimensional. Podem-se fazer assim *queries* sobre esta estrutura, com vista a detectar tendências e a tirar conclusões. Pesquisas e questões que nas bases de dados tradicionais demorariam três dias a ser respondidas podem ser solucionadas em apenas três segundos na abordagem OLAP.

Figura 4.4
Conceito da
análise OLAP



Podem ser distinguidos essencialmente três tipos de OLAP: o *relational* OLAP (ROLAP), o *multidimensional* OLAP (MOLAP) e o *hybrid* OLAP (HOLAP). Vejamos em pormenor as características de cada um deles.

O ROLAP

O ROLAP caracteriza-se por manter os dados nas tabelas relacionais originais, ao mesmo tempo que gera outras em que vai armazenar os valores agregados. Estes dados agregados são somas com um baixo nível de detalhe derivadas dos dados. Um exemplo natural é a agregação de valores respeitantes às várias unidades de tempo (dia, semana, mês, ano).

Esta capacidade de aglutinar dados permite aumentar significativamente os desempenhos, tendo em conta que muitos dos valores solicitados pelas interrogações são pré-calculados. Apesar disto, o ROLAP é uma solução globalmente mais lenta do que outras opções, dado que é o modelo relacional que tem de sustentar o trajecto de todas as uniões entre tabelas. Esta desvantagem pode ser compensada pela sua pequena dimensão.

O MOLAP

O MOLAP utiliza uma estrutura de dados multidimensional para armazenar a informação. A estrutura é gerida por um motor de base de dados multidimensional e não por um motor relacional. As agregações são calculadas automaticamente como parte da estrutura multidimensional. O MOLAP é extremamente rápido em termos de resposta às questões dos utilizadores, mas, infelizmente, apresenta desvantagens a nível do espaço ocupado e do tempo dispendido na sua criação.

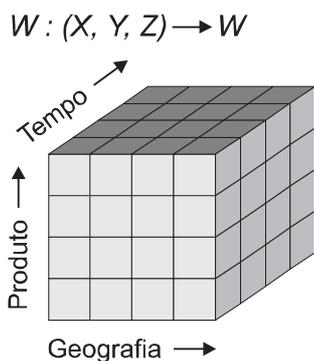
O HOLAP

O HOLAP é um meio-termo entre as duas tecnologias anteriores, assumindo aspectos e atributos de ambas. Este modelo deixa os dados nas tabelas relacionais e guarda as agregações como uma estrutura multidimensional. O HOLAP assume-se assim como uma opção de compromisso, quer em termos de rapidez de resposta, quer em termos da dimensão.

CUBOS OLAP

Nas teorias de base de dados, os cubos OLAP são representações abstractas de uma projecção de uma relação RDBMS (*relational database management system*). Podem ser descritos, em termos de uma função, da seguinte forma:

Figura 4.5
Cubo de
informação
para análise OLAP



Nesta função, X, Y e Z são as chaves e representam eixos e W o valor que resulta do seu cruzamento e preenche a célula respectiva do cubo.

O cubo OLAP, criado a partir de um esquema em estrela de tabelas, permite formas de navegação fáceis e intuitivas, através de dife-

rentes graus de detalhe da informação. Através de uma funcionalidade denominada *drill*, um utilizador final pode navegar nos dados que interessam à sua análise, aumentando o grau de detalhe (*drill down*) e descendo, consequentemente, a dados cada vez mais operacionais, ou diminuindo o detalhe (*drill up*), subindo a uma informação cada vez mais agregada.

De realçar, também, a sempre útil funcionalidade apresentada pelas ferramentas OLAP de rearranjar as linhas e colunas de dados em análise, o chamado *slice and dice*. O utilizador pode decidir cruzar de forma diversa as várias dimensões ao seu dispor (e. g., vendas x área geográfica x intervalo de tempo ou área geográfica x categoria de produto x responsável comercial), trocando simplesmente a ordem das colunas e linhas e suprimir ou visualizar algumas delas.

Figura 4.6
Imagem de ecrã
da aplicação
Analysis Studio

The screenshot shows the Analysis Studio interface with a pivot table. The 'Insertable Objects' pane on the left lists 'SalesCube (1 of 7)', 'Dates', 'Products (5)', 'Locations', 'Branch', 'Order Method', 'Margin Range', and 'Measures'. The main area displays a pivot table with 'Revenue' as the row label, '2004' and '2005' as column labels, and 'Dates' as the context label. The table data is as follows:

Revenue	2004	2005	Dates
Consumer Electronics	\$38,908,096.24	\$163,003,646.58	\$201,911,742.82
Home Theatre	\$38,468,425.88	\$67,685,103.86	\$106,153,529.75
Entertainment Media	\$158,854.12	\$358,203.17	\$517,057.29
Lifestyle Products	\$186,606.39	\$152,019.02	\$338,625.41
Home Office	\$396,093.00	\$116,406.00	\$512,499.00
Products	\$78,118,075.64	\$231,315,378.63	\$309,433,454.27
E-mail	\$17,155,079.76	\$45,176,068.23	\$62,331,147.99
Fax	\$10,254,519.41	\$30,201,548.08	\$40,456,067.49
Mail	\$16,383,665.13	\$50,084,341.71	\$66,468,006.84
Sales visit	\$6,484,546.71	\$22,026,678.62	\$28,511,225.32
Special	\$1,464,301.80	\$3,394,486.80	\$4,858,788.60
Telephone	\$15,837,608.10	\$50,088,887.36	\$65,926,495.47
Web	\$10,538,354.73	\$30,343,367.83	\$40,881,722.56
Order Method	\$78,118,075.64	\$231,315,378.63	\$309,433,454.27

APLICAÇÕES DA TECNOLOGIA OLAP – PROJECTOS ANALÍTICOS DEPARTAMENTAIS

A realidade do negócio das empresas actuais pressupõe, de uma forma transversal a quase todos os sectores e áreas de actividade, um

conjunto de áreas departamentais ou funcionais de cuja análise dos dados depende a consolidação de conclusões e conseqüentes decisões. Estes projectos, que são muitas vezes concebidos de forma gradual e posteriormente integrados, são o cerne de um ambiente de BI escalável, completo e sem redundâncias.

Neste subcapítulo, elencamos cinco áreas funcionais, em que identificamos nove projectos específicos de cubos multidimensionais (OLAP). Apesar das especificidades inerentes a cada organização, as áreas, dimensões e respectivas medidas (colunas de dados) são comuns à maioria das empresas produtoras de bens e/ou prestadoras de serviços.

Finanças

Nas tarefas inerentes à gestão financeira, é importante que os decisores tenham um retrato objectivo, em tempo real, da situação patrimonial, do desempenho das várias áreas de negócio e/ou centros de custo e dos movimentos de tesouraria que ocorreram, efectivamente, em determinado período. As fontes de dados são, preferencialmente, ERP e ficheiros Microsoft Excel.

Figura 4.7
Balço
multidimensional

TEMPO	RUBRICAS DO BALANÇO	ORGANIZAÇÃO	% PLANEADO	INDICADORES
Anos	Activo imobilizado	Grupo (<i>Corporate</i>)	>120%	Valores actuais
Trimestres	Activo circulante	Empresa	110-120%	Valores orçamentados
Meses	Passivo c/ prazo	Divisão	100-110%	<i>Rolling forecast</i>
	Passivo m/l prazo		90-100%	Valores de abertura
	[...]			Valores de fecho

Figura 4.8
Análise
financeira

TEMPO	ORGANIZAÇÃO	KPI FINANCEIROS	INDICADORES
Anos	Grupo (<i>Corporate</i>)	EPS	Actual
Trimestres	Empresa	EBITDA	Planeado
Meses	Divisão	Rendibilidade vendas	<i>Rolling forecast</i>
		Rendibilidade activos	
		Rendibilidade c. próprios	
		Grau alavanca financeira	
		VAL	
		TIR	

Figura 4.9
Análise de
cash-flows

TEMPO	ORGANIZAÇÃO	CASH FLOWS	INDICADORES
Anos	Grupo (<i>Corporate</i>)	Origens de fundos	Variação líquida (act)
Trimestres	Empresa	– Área operacional	Variação líquida (plan)
Meses	Divisão	– Área investimento	Variação líquida (RF)
	[...]	Utilização dos fundos	
		– Pagamento débitos	
		– Investimento	

Comercial/ Vendas

Na análise comercial de uma organização, é importante considerar as suas vendas numa perspectiva multidimensional (região, canal de venda, categorias de produtos, entre outros), discernindo tendências e tirando conclusões. Importa compreender os segmentos ou nichos específicos de maior rentabilidade e, com isso, preparar decisões relativamente a *pricing*, política de distribuição e políticas de incentivos comerciais. As fontes de dados são, essencialmente, ERP e aplicações específicas de CRM.

Figura 4.10
Análise
de vendas

TEMPO	ORGANIZAÇÃO	C. VENDA	PRODUTOS	CLIENTES	INDICADORES
Anos	Grupo (<i>Corporate</i>)	Canal 1	Linhas produtos	Segmento 1	Unidades vendidas
Trimestres	Empresa	Canal 2	Marcas	Segmento 2	Receitas
Meses	Divisão	Canal 3	Produtos	Segmento 3	% Desconto
Semanas	Gestores comerciais	[...]	SKU (*)	[...]	% Comissões
Ano n-1					% Reclamações
[...]					Preço médio venda

(*) *Stock keeping unit*

Figura 4.11
Rentabilidade
de produto/cliente

TEMPO	ORGANIZAÇÃO	C. VENDA	PRODUTOS	CLIENTES	INDICADORES
Anos	Grupo (<i>Corporate</i>)	Canal 1	Linhas produtos	Segmento 1	Custo médio
Trimestres	Empresa	Canal 2	Marcas	Segmento 2	– C. m. produção
Meses	Divisão	Canal 3	Produtos	Segmento 3	– C. m. distribuição
Semanas	Gestores comerciais	[...]	SKU	[...]	Lucro bruto
Ano n-1					Lucro líquido
[...]					

Marketing

Na análise de *marketing*, é importante ter uma perspectiva «macro», mais estratégica, do mercado que vamos abordar: dimensão, segmentação explícita ou implícita, vantagens/desvantagens de determinado posicionamento. Numa perspectiva «micro», mais operacional, importa conhecer o impacto das campanhas de publicidade, de alterações pontuais no *marketing-mix* dos produtos (*product*, *price*, *place* e *promotion*).

Figura 4.12
Análise de
marketing
estratégico

Os ERP, aplicações de CRM, bases de dados com informação diversa de *marketing research* são as fontes de informação mais usuais.

TEMPO	CANAIS	PRODUTOS	CLIENTES	INDICADORES
Anos	Canal 1	Linhas produtos	Mercado Global	Número clientes
Trimestres	Canal 2	Marcas	– Segmento 1	LTV (*) médio clientes
Meses	Canal 3		– Segmento 2	
Semanas	[...]		– Segmento 3	
Ano n-1				
[...]				

(*) *Life Time Value*

Figura 4.13
Análise de
marketing
operacional

TEMPO	CAMPANHAS	MENSAGENS	SEG. MERC.	PRODUTOS	INDICADORES
Anos	Campanha 1	Linhas produtos	Merc. global	Linhas produtos	Número clientes
Trimestres	Campanha 2	Marcas	– Segmento 1	Marcas	Número produtos
Meses	Canal 3	Produtos	– Segmento 2	Produtos	Proveitos/Campanha
Semanas	[...]		– Segmento 3		Custos/Campanha
Ano n-1					% Respostas
[...]					

Produção

Na produção, os decisores tentam otimizar a correlação entre diversas variáveis críticas (necessidades de produção, capacidade instalada, matérias-primas/subsidiárias disponíveis, recursos necessários, custo da produção), de forma maximizar as vendas e otimizar a capacidade instalada. As aplicações de *forecasting* de vendas, de gestão da produção e da SCM (*supply chain management*) são importantes fontes de dados.

Figura 4.14
Gestão
da capacidade
de produção

TEMPO	PRODUTOS	ETAPAS PRODUÇÃO	LINHAS PROD.	INDICADORES
Anos	Linhas produtos	Fabricação	Fábrica 1	Unidades planeadas
Trimestres	Marcas	Montagem	– Linha 1	Unidades produzidas
Meses	Produtos	Inspeção	– Linha 2	% defeitos
Semanas	SKU	Embalagem	– Linha 3	<i>Uptime</i>
Ano n-1		[...]	Fábrica 2	<i>Downtime</i>
[...]			[...]	% Capacidade utilizada

Gestão de recursos humanos

Como prioridade, é necessário que a função de recursos humanos apresente um conhecimento objectivo das suas características demográficas e, a partir daí, permita compreender a sua adequação às competências necessárias à organização. De forma complementar, torna-se também importante compreender as competências fundamentais, necessidades formativas e a avaliação do desempenho. Os módulos de gestão de pessoal dos ERP e ficheiros Excel com informação diversa serão as fontes que alimentarão estes cubos.

Figura 4.15
Administração
de recursos
humanos

TEMPO	ORGANIZAÇÃO	GRUPO PROF.	CONTRATO	AVAL.	INDICADORES
Anos	Grupo (<i>Corporate</i>)	Administração	<i>Full time</i>	1	Núm. empregados
Trimestres	Empresa	Directores 1.ª linha	– Efectivo	2	FTE
Meses	Divisão	Chefes de sector	– Prazo	3	Custos totais RH
Ano n-1		Administ./Comer.	– Trab. temp.	4	% Incentivos
[...]		[...]		5	% Promoções
					% Formação

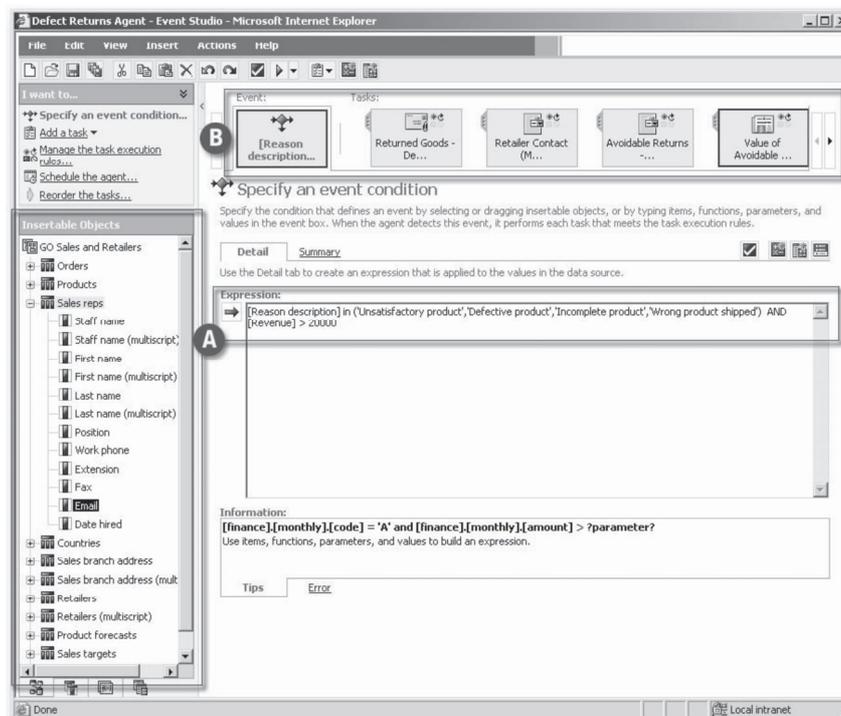
APLICAÇÕES DE *EVENT MANAGEMENT*

Este tipo de aplicações é orientado para uma monitorização contínua dos processos de negócio sem interacção do utilizador. Funciona através do acompanhamento de eventos que necessitam de ser vigiados, utilizando uma automatização de processos de negócio e de deci-

são, de forma a reduzir drasticamente o tempo de passagem à acção e respectiva resolução. Um exemplo pode ser um alerta automático para um utilizador quando as vendas desçam abaixo de um determinado valor ou até mesmo a mudança de um valor numa base de dados (como um preço de um produto) quando o seu *stock* baixar a um determinado número de unidades.

Estas ferramentas funcionam através da definição, por parte dos utilizadores, de condições e agentes em que são especificados os eventos a ser monitorizados e as respectivas tarefas a ser executadas caso as condições sejam cumpridas de acordo com o contexto em que se inserem.

Figura 4.16
Imagem de ecrã
da aplicação de
Cognos
Business Event
Management®



Estas ferramentas estão a ser aprimoradas com o advento da tecnologia BAM, de que falaremos num outro capítulo.

FERRAMENTAS DE VISUALIZAÇÃO: DASHBOARDING E SCORECARDING

Para fáceis monitorização, compreensão e assimilação da informação é essencial que haja interfaces e aplicações de *front-end* ao serviço

dos gestores e decisores. A visualização é, claramente, um dos factores fundamentais à assimilação e percepção de quem decide. A análise tabular, numérica e tradicional já não é assertiva face às necessidades de análise e consolidação de contextos caracterizados por diversas variáveis geográficas, demográficas, comerciais e organizacionais.

Desde os já «pré-históricos» *tableaux de bord*, disponibilizados em ambientes pouco simpáticos, há 20 anos, toda a tecnologia de visualização de dados teve enormes desenvolvimentos, impensáveis no passado recente. Hoje em dia, utilizam-se, de forma generalizada, os gráficos e tabelas mais básicos, indicadores e sinaléticas, animações, ferramentas interactivas de análise multidimensional e as chamadas aplicações de *scorecarding* e *dashboarding*, por todos aqueles que necessitam de conhecimento fácil e oportuno, «à distância de um clique».

A introdução de ferramentas modernas de visualização é pertinente em inúmeras tarefas de análise e decisão actuais:

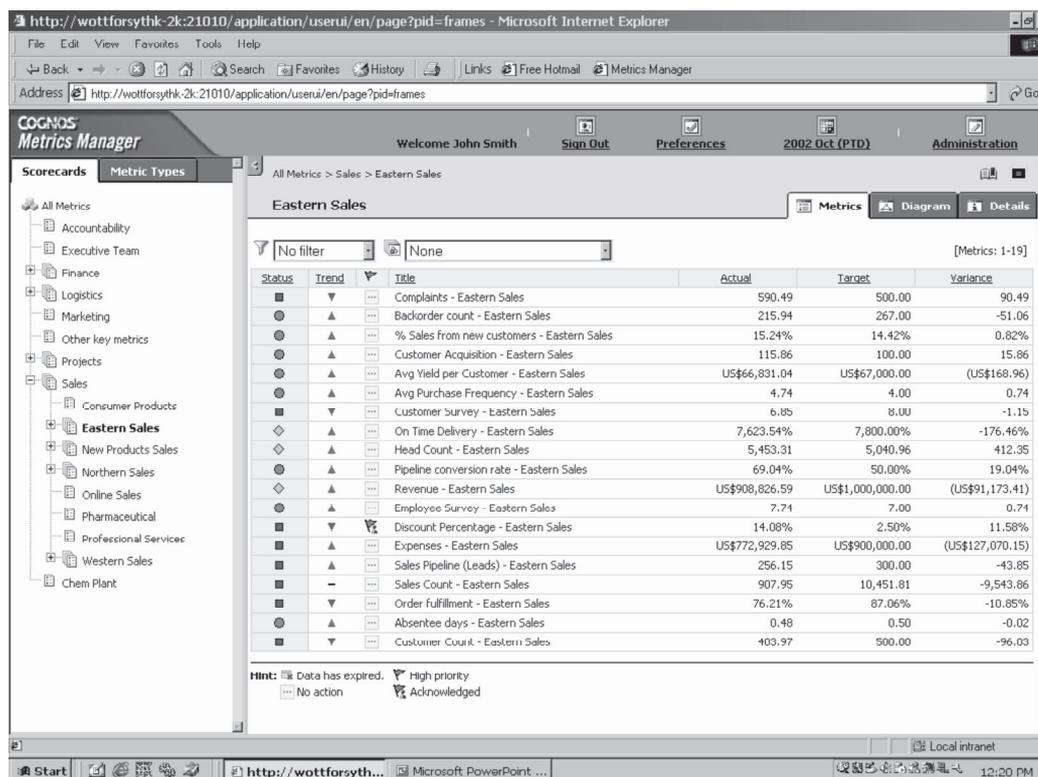
- na gestão de produção, identificando volumes, capacidades de produção utilizadas e rendibilidade;
- na gestão comercial, visualizando volumes, margens e segmentação das vendas por múltiplos critérios, tendo em vista a identificação de clientes e nichos de mercado que maximizem o lucro;
- na gestão de *marketing*, analisando de modo «geoespacial» as características demográficas dos mercados em que a empresa está inserida;
- na gestão de recursos humanos, identificando e diagnosticando o estado da formação e das competências e tendências de *turnover* nas várias áreas da organização;
- na gestão logística, visualizando inventários e requisitos das cadeias de aprovisionamento, de forma a otimizar compras e níveis de existências.

AS APLICAÇÕES DE SCORECARDING

Os *scorecards* aperfeiçoam o *reporting* tradicional ao utilizarem medidas-alvo pré-definidas (*e. g.*, objectivos de vendas, razões de rendibilidade e eficiência) com os respectivos dados actuais e desvios. Baseando-se em informação diversa obtida nos sistemas operacionais, os *scorecards* agregam dados diversos em indicadores-chave, representativos do desempenho. Podem ser utilizados em sistemas de controlo operacional para períodos determinados (como no caso dos

projectos de Six Sigma, de aperfeiçoamento do desempenho), bem como em EIS (*executive information systems*), contendo informação composta e/ou agregada, de conteúdo estratégico (de que é exemplo a metodologia do *balanced scorecard*).

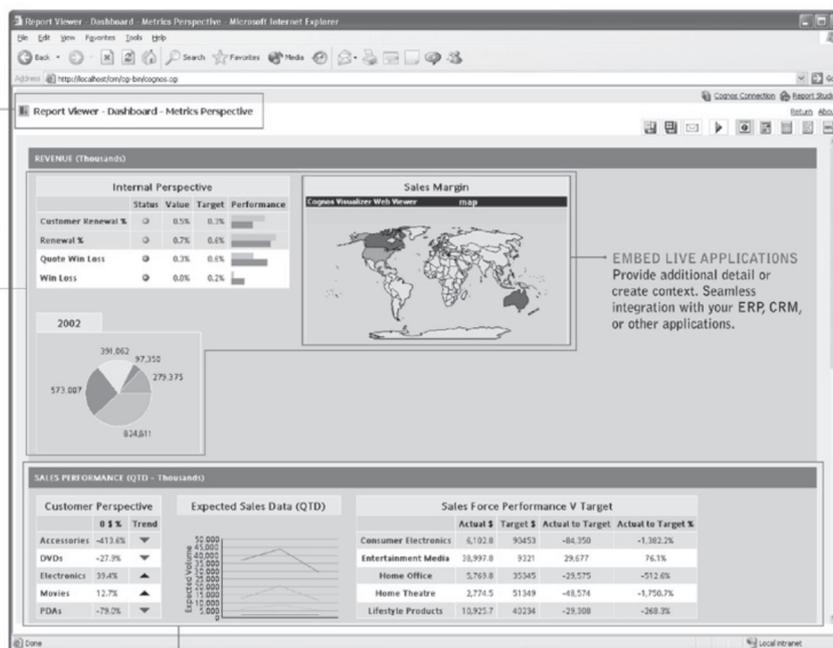
Figura 4.17
Imagem de ecrã
da aplicação
de *scorecarding*
Cognos Metrics
Manager®



O DASHBOARDING

Os *dashboards* apresentam como principal vantagem a capacidade de comunicarem quantidades complexas de informação de uma forma rápida e intuitiva. Mais concretamente, traduzem os dados recebidos dos sistemas operacionais de uma forma graficamente apelativa e amigável do utilizador, através de mapas, gráficos diversos, tabelas, entre outros. Na actualidade, é normal haver *dashboards* dinâmicos, que permitem efectuar operações de *drill through* (pesquisa aprofundada) noutras fontes de dados para ver com maior detalhe os dados subjacentes ao que os *dashboards* nos revela.

Figura 4.18
Imagem de ecrã
da aplicação de
dashboarding
Cognos
Reportnet®



Em síntese, as capacidades das tecnologias de *dashboarding* e *scorecarding* são essencialmente complementares nos processos e tarefas de análise e decisão empresariais, como é explicitado abaixo:

Figura 4.19
Quadro-síntese
das
potencialidades
e dos aspectos
diferenciadores
das tecnologias
de *dashboarding*
e *scorecarding*



A VISUALIZAÇÃO DE INFORMAÇÃO MULTIDIMENSIONAL

A especificidade dos dados e da informação disponibilizada pelas ferramentas de *business intelligence* (transversal à organização, multifuncional e multidimensional), obriga, frequentemente, à existência de aplicações mais sofisticadas de visualização e análise. Com efeito, discernir factos e tendências com inclusão de múltiplas variá-

veis (quantidade, importância, tempo, geográficas, entre outras) pressupõe outras capacidades para além das normalmente existentes num simples gráfico de duas dimensões.

Genericamente, estas ferramentas devem permitir:

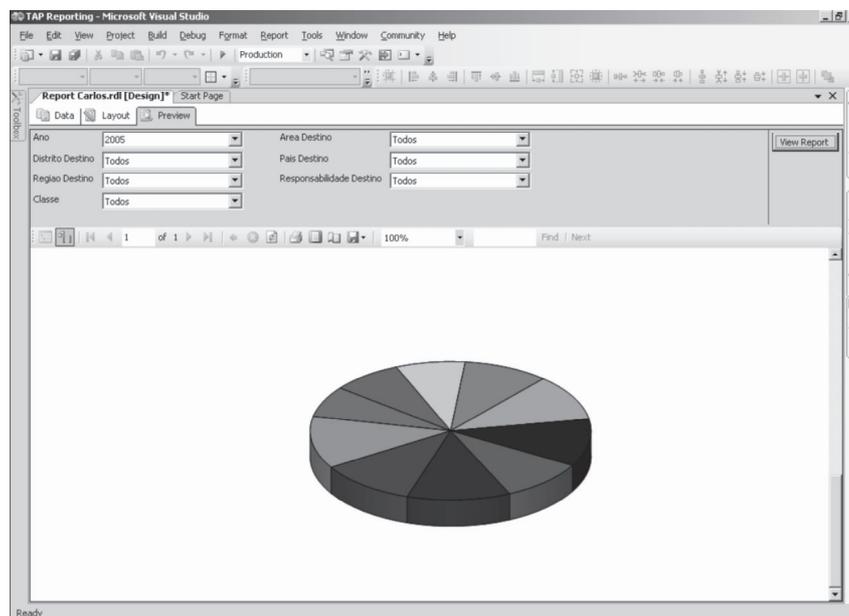
- a visualização de várias dimensões e variáveis;
- a interactividade com os utilizadores, de modo a permitir uma visualização de vários ângulos e enquadramentos e que potencie a segmentação da informação em parcelas (tipo *slice and dice*).

Existem várias ferramentas para visualização de dados multidimensionais. Apresentamos algumas em seguida.

Pie charts multidimensionais

Este grafismo permite normalmente analisar até três variáveis dependentes (correspondentes ao tamanho, altura e cor), a que correspondem três variáveis independentes (a fatia, linha e coluna), permitindo uma avaliação rápida e intuitiva dos dados.

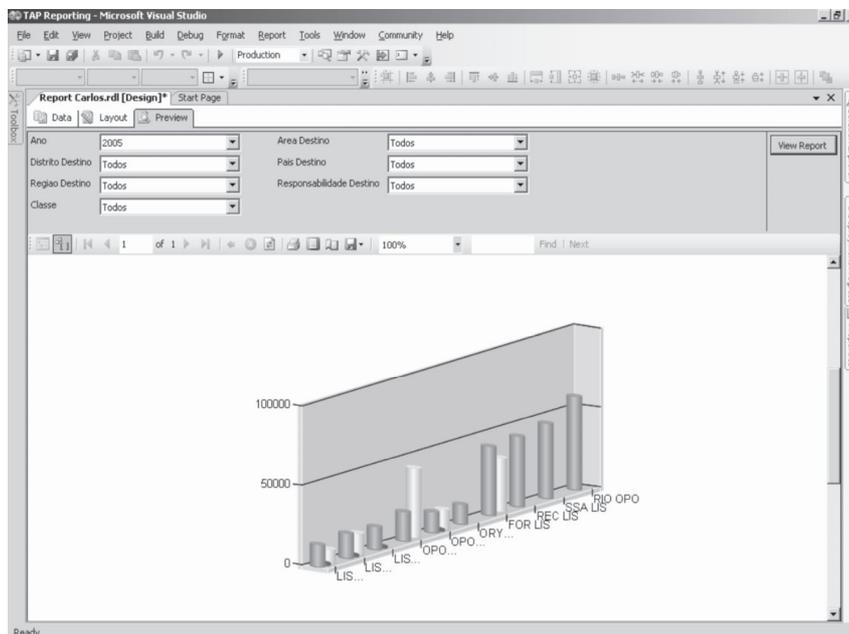
Figura 4.20
Exemplo
de um *pie chart*
multidimensional



Gráficos de barras multidimensionais

Os gráficos de barras multidimensionais permitem aos utilizadores (através dos eixos e dos níveis de cor) a visualização de dados que assentam em diversas variáveis.

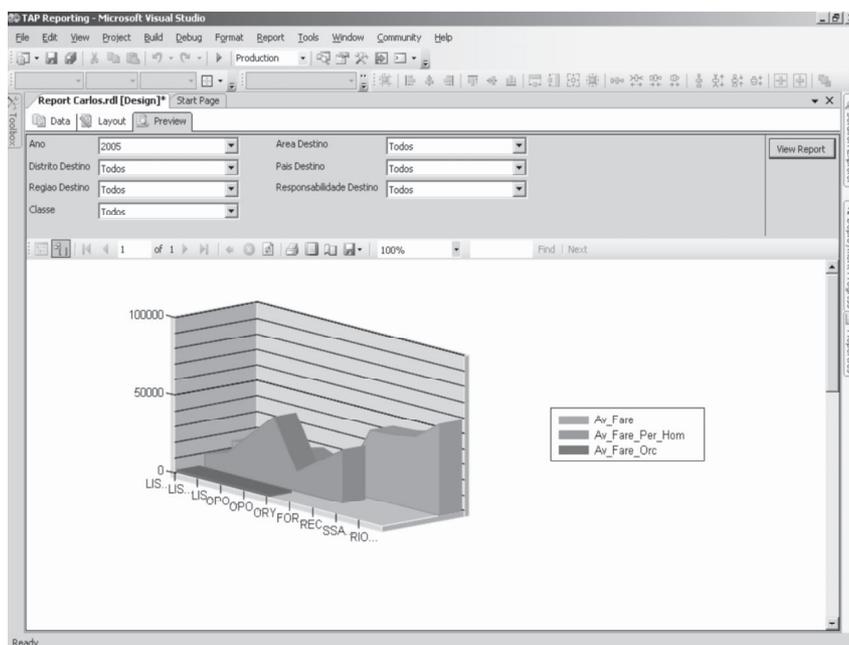
Figura 4.21
Exemplo de um gráfico de barras multidimensional



Histogramas multidimensionais

Os histogramas, que representam várias dimensões, estão mais vocacionados para análises de grandes volumes de dados e em que seja pertinente avaliar a sua densidade de distribuição.

Figura 4.22
Exemplo de um histograma multidimensional



Aplicações avançadas

Surgiram recentemente ferramentas com capacidades avançadas que proporcionam aos utilizadores diferentes experiências de visualização e percepção.

Estas funcionalidades procuram potenciar uma rápida assimilação cognitiva da realidade.

As animações implicam facilidade, rapidez e eficácia no discernimento da tendência e na velocidade do seu desenvolvimento.

Já as visualizações em «imersão», tal como num jogo de vídeo de realidade virtual, colocam o utilizador no meio de um cenário que contém a representação multidimensional dos dados. Utilizadas inicialmente nas áreas de inovação e desenvolvimento do produto, são já aplicadas para uma contextualização do decisor em cenários complexos.

Por último, e directamente relacionadas com as anteriores, estão as capacidades de interactividade accionadas pelo utilizador. Enquanto as aplicações mais tradicionais apresentam barras ou linhas estáticas, estas novas possibilidades oferecem, por exemplo, a possibilidade de «rodar» gráficos e imagens, de forma a mostrar conhecimento oculto, filtrar dados acima ou abaixo de um determinado valor ou efectuar *zoom* em determinadas partes das imagens. Esta interacção directa poupa aos utilizadores tempo gasto em *queries* e a formatação do *reporting*, que pode assim ser utilizado na análise da informação.

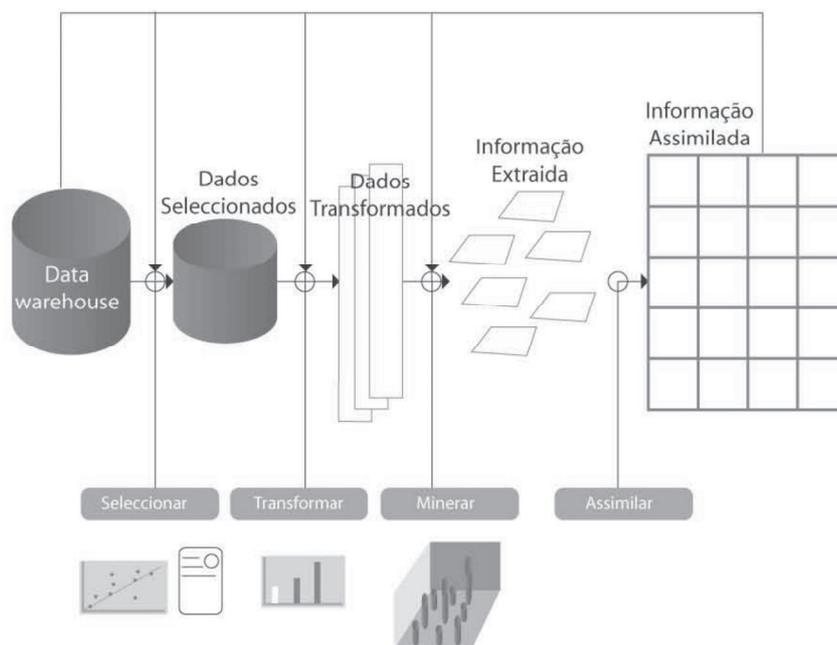
DATA MINING

A *data mining* ou a «mineração de dados» é uma das ferramentas de BI mais conhecidas e com utilização mais generalizada no mundo empresarial.

***Data mining* – conceito que engloba todos os processos que, através de uma diversidade de ferramentas tecnológicas de análise, permitem descobrir padrões e relações num determinado conjunto de dados.**

As aplicações desta abordagem são essencialmente descritivas (retrata uma realidade actual específica) e predictivas (permite fazer previsões e conclusões através da concepção de modelos, tendo em conta os padrões actuais de comportamento detectados).

Figura 4.23
Representação
gráfica
da arquitectura
de um sistema
de *data mining*



As aplicações empresariais de um sistema de *data mining* são muitas e variadas, mas podem ser agregadas em dois grandes grupos: a optimização da gestão da estratégia comercial/*marketing* – através da análise do mercado, dos segmentos e clientes – e a procura de eficiência nos custos – através da racionalização dos diversos *inputs* (materiais e humanos) de cada processo produtivo.

De realçar que a *data mining* não substitui o factor humano na análise dos dados. O facto de permitir definir padrões e tendências não dá, automaticamente, ao gestor o valor e impacto dessas realidades para o negócio.

DATA MINING DESCRITIVA

Quando falamos em *data mining* descritiva referimo-nos à descrição dos seus principais atributos estatísticos (valores médios, desvio-padrão e desvio médio, entre outros) e da visualização destes valores (através de gráficos e tabelas). É igualmente importante descortinar, nesta análise, ligações entre as diversas variáveis (*e. g.*, vislumbrar dados que ocorram simultaneamente).

DATA MINING PREDICTIVA

Uma *data mining* meramente descritiva tem um valor limitado para o entendimento do negócio de cada organização. Permite, com efeito,

analisar com objectividade os valores e tendências históricas mas não traz grande capacidade prospectiva para apoiar decisões. É, pois, essencial construir modelos predictivos baseados nos dados apurados. Esses modelos (baseados em algoritmos) devem ser elaborados, testados e validados face à realidade, com vista a apurar a sua fiabilidade. Este será um processo de natureza cíclica, em permanente revisão, tendo em conta a constante mudança das variáveis de negócio.

O PROCESSO DE *DATA MINING*

Um projecto de *data mining* deverá ser prosseguido através das seguintes etapas:

- Definir o *business case*

É essencial definir de forma clara o objectivo e as metas do projecto, para explicitar o impacto que se pode esperar no negócio e justificar o investimento subjacente. Esta definição inicial deverá, pois, incluir a explicação da maneira como a descoberta de conhecimento pertinente (os designados «projectos de *knowledge discovery*») através de *data mining* afectará o negócio. Uma definição de finalidade será, por exemplo, a minimização do risco de crédito nos produtos de crédito ao consumo de um determinado banco comercial – o que originaria uma pesquisa ao histórico do perfil de clientes não cumpridores.

- Construir a base de dados-alvo da *data mining*

Esta é uma das etapas fundamentais do projecto e uma das que consomem mais tempo. Os dados a ser explorados deverão ser agrupados numa base de dados própria (separada do DW corporativo), com vista a uma exploração que tenha em conta a necessidade de constantes iterações e os devidos cuidados na gestão dos recursos informáticos. Os passos/tarefas a serem executadas são idênticos aos já descritos na construção do DW: recolha, descrição e selecção dos dados, análise de *data quality* e respectiva limpeza, consolidação e integração, construção dos metadados, carregamento e os necessários trabalhos de manutenção da base de dados.

- Explorar os dados

Na fase de exploração, tentamos identificar os campos (variáveis) mais importantes na análise de relação ou predição e determinar que valores derivados serão mais úteis. Tendo em conta que estaremos a lidar com conjuntos de dados com centenas ou mesmo milhares de colunas, a capacidade de processamento e resposta da aplicação informática de suporte é um factor fundamental.

- Preparar os dados para a modelização

Este passo, o último antes da construção dos modelos, pressupõe, inicialmente, a selecção das variáveis e das colunas – uma vez que, acumulando toda a informação disponível sem um critério selectivo, poderemos chegar a um modelo incorrecto, com variáveis irrelevantes. Terá ainda de ser encarada a necessidade de construir e/ou transformar variáveis (*e. g.*, de valores absolutos para razões mais compostas) tendo em vista o aumento da fiabilidade do modelo e o seu enquadramento nos pressupostos de suporte à decisão.

- Construir um protótipo do modelo

A construção do modelo é feita por via de múltiplas iterações, explorando vias alternativas (*train & test*), com vista a avaliar quais são as mais indicadas para resolver o *business case*. A decisão relativamente ao tipo de modelo também é importante – seja, por exemplo, uma árvore de decisão, um *neural net* ou uma regressão logística. A decisão irá afectar o tempo e as condições de preparação subjacentes;

- Avaliar o modelo

O modelo proposto terá de ser avaliado pelos seus resultados e pelo seu grau de importância, detectando eventuais erros de construção. É particularmente importante fazer a sua validação externa, isto é, atestar se o modelo concebido reflecte o mundo real e se é pertinente para as explicações que se procuram – uma vez que, mesmo estando correcto, poderá não ser o mais adequado;

- Implementar o modelo validado

Operacionalizar o modelo é, pois, o último passo. A sua implementação deve ser efectuada após uma correcta validação interna e externa, disseminada pelos utilizadores-chave (analistas e decisores) do processo de negócio respectivo e seguida por uma monitorização constante dos padrões que presidiram à sua elaboração.

TESTE OS SEUS CONHECIMENTOS

1. Quais são as vantagens mais importantes trazidas pelas aplicações OLAP na análise da informação do negócio?
2. Quais são as principais aplicações das tecnologias de *dashboarding* e *scorecarding*?
3. Explícite os principais requisitos e etapas para a construção de um projecto de *data mining*.

