#### **GUSTAVO GATTASS AYUB**

INTEGRAÇÃO DE DADOS PARA A INTELIGÊNCIA EMPRESARIAL
EM TEMPO REAL

#### **GUSTAVO GATTASS AYUB**

# INTEGRAÇÃO DE DADOS PARA A INTELIGÊNCIA EMPRESARIAL EM TEMPO REAL

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do Título de Mestre em Engenharia Elétrica.

São Paulo 2011

#### **GUSTAVO GATTASS AYUB**

# INTEGRAÇÃO DE DADOS PARA A INTELIGÊNCIA EMPRESARIAL EM TEMPO REAL

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do Título de Mestre em Engenharia Elétrica.

Área de Concentração: Sistemas Digitais

Orientador:

Prof. Dr. Jorge Rady de Almeida Junior

São Paulo 2011 Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.		
São Paulo, de maio de 2011.		
Assinatura do autor		
Assinatura do orientador		

#### FICHA CATALOGRÁFICA

**Ayub, Gustavo Gattass** 

Integração de dados para a inteligência empresarial em tempo real / G.G. Ayub -- São Paulo, 2011. 113 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1. Inteligência empresarial 2. Banco de dados ativos 3. Tempo real 4. Depósito de dados I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II. t.

#### **AGRADECIMENTOS**

Ao Prof. Dr. Jorge Rady de Almeida Junior, pela atenção e apoio durante o processo de definição e orientação.

À Escola Politécnica da Universidade de São Paulo, pela oportunidade de realização do curso de mestrado.

**RESUMO** 

Nos últimos anos, a utilização do Data Warehouse mudou significativamente. Nos

dias de hoje, as empresas precisam distribuir informações, num intervalo de tempo

adequado, para os colaboradores que atuam na linha de frente do negócio, de modo

que estes possam obter um desempenho superior na execução de suas tarefas.

O presente trabalho apresenta as novas tendências em inteligência empresarial,

dando destaque aos aspectos táticos e operacionais, a necessidade de redução na

latência de atualização dos dados, as limitações do ambiente de Data Warehouse

nesse contexto. O desenvolvimento deste trabalho baseou-se na revisão da

literatura existente, no estudo de novos mecanismos de integração de dados, na

apresentação de casos reais de aplicação e em um relato de caso prático.

Este trabalho buscou destacar o papel fundamental que as novas tecnologias de

integração de dados têm na criação de novos sistemas de apoio e automação de

decisões nas empresas realizando negócios em tempo real.

Palavras-Chave: Inteligência empresarial, Banco de Dados Ativos, Tempo Real,

Depósito de Dados

vi

vii

**ABSTRACT** 

During the last years many things have changed in the utilization the Data

Warehouse. Nowdays, enterprises need to distribute information in a timely fashion

to front-line workers, enabling them to deliver superior performance in the execution

of their duties.

This work presents new trends in business intelligence, emphasizing the tactical and

operational aspects, the demand for reducing data latency, the limitations of the Data

Warehouse in this context. The development of this work was based on the review of

existing literature, the study of new mechanisms for data integration, the presentation

of real case scenarios of application and a practical case story.

This study highlights the key role that the new data integration technologies have

while are creating new automating decisions systems in companies doing business in

real time.

**Keywords:** Business Intelligence, Active Databases, Real-Time, Data Warehouse.

# SUMÁRIO

1	INT	FRODUÇÃO	
	1.1	JUSTIFICATIVA	2
	1.2	Objetivos	3
	1.3	ESTRUTURA DO TRABALHO	4
2	ΑII	NTELIGÊNCIA EMPRESARIAL EM TEMPO REAL	5
	2.1	REVISÃO DA LITERATURA	9
	2.1	1.1 Active Data Warehousing	
	2.1	1.2 Business Activity Monitoring	14
	2.1	1.3 Business Performance Management	20
	2.1	1.4 Vigilant Information Systems	27
3	TE	NDÊNCIAS EM INTELIGÊNCIA EMPRESARIAL	30
	3.1	ENFOQUE OPERACIONAL	
	3.2	REDUÇÃO DAS LATÊNCIAS	
	3.3	DECISÕES EM TEMPO REAL	35
	3.4	Novas Fontes de Dados	
	3.5	DESAFIOS TECNOLÓGICOS	
	3.6	NÍVEIS DE MATURIDADE	
	3.6		
	3.6	3	42
	3.6	3	
	3.6	3	
	3.7		
	3.7		
	3.7	7.2 O DW 2.0	47
4	INT	TEGRAÇÃO DE DADOS EM TEMPO REAL	49
	4.1	Captura ou Aquisição dos Dados	50
	4.1	1.1 Mini Batches ou Micro Batches	51
	4.1	1.2 Change Data Capture	52
	4.1	1.3 Trickle Feed	53
	4.2	ARMAZENAMENTO E PROCESSAMENTO DOS DADOS	53
	4.2	P.1 Operational Data Stores	53
	4.2	2.2 Enterprise Information Integration	55
	4.2	2.3 In-Memory Databases	56
	4.2	2.4 Stream Processing	57
	4.3	VISUALIZAÇÃO DE DADOS	59
	4.3	3.1 Dashboards	59
	4.3	3.2 Sistemas de Informação Geográfica	63
	4.4	CASOS REAIS DE APLICAÇÃO	64
	4.4	1.1 Quicken Loans	64
	4.4	1.2 Yahoo!	66
	4.5	Considerações	69

5	AA	DOÇÃO DA INTELIGÊNCIA EMPRESARIAL EM TEMPO REAL	70
	5.1	REDUZINDO AS LATÊNCIAS	70
	5.2	O MODELO DE IDENTIFICAÇÃO E JUSTIFICAÇÃO DE MCGEE	71
	5.3	A VISÃO DO PROJETO	73
	5.4	OS INVESTIMENTOS NECESSÁRIOS NA INFRAESTRUTURA DE TECNOLOGIA DA INFORMAÇÃO	75
	5.5	Os Desafios Técnicos	
	5.5.	- 4	
	5.5.	· · · · · · · · · · · · · · · · · · ·	
	5.6	SELECIONAR AS TECNOLOGIAS ADEQUADAS	
	5.6.	,	
	5.7	A CAPTURA E AQUISIÇÃO DOS DADOS	
	5.8	O ARMAZENAMENTO E PROCESSAMENTO DOS DADOS	
	5.8.		
	5.8.		
	5.8.		
	5.8.		
	5.8.	, ,	
	5.8.	ŭ	
	5.8.	· · · · · · · · · · · · · · · · · · ·	
	5.8.	,	
	5.9	Considerações	92
6	REI	LATO DE CASO	93
	6.1	O SISTEMA BAM (BUSINESS ACTIVITY MONITORING)	94
	6.2	O PROBLEMA DA LATÊNCIA NOS DADOS	
	6.3	REQUISITOS PARA A LATÊNCIA NOS DADOS	96
	6.4	ALTERNATIVAS PARA A REDUÇÃO DAS LATÊNCIAS	
	6.5	CICLO DE VIDA DOS DADOS	
	6.6	PROVA DE CONCEITO	101
	6.7	RESULTADOS OBTIDOS	102
7	CO	NCLUSÕES	103
	7.1	RESULTADOS OBTIDOS	103
	7.2	TRABALHOS FUTUROS	104
8	REI	FERÊNCIAS BIBLIOGRÁFICAS	105

# LISTA DE ILUSTRAÇÕES

FIGURA 1 - UM MODELO PARA O <i>BUSINESS INTELLIGENCE</i> (WATSON; WIXOM, 2007)	7
FIGURA 2 - O CICLO DO BPM (ECKERSON, 2006)	22
FIGURA 3 - O PROCESSO DECISÓRIO (GOLFARELLI; RIZZI; CELLA, 2004)	25
FIGURA 4 - VIGILANT INFORMATION SYSTEMS E OS DEMAIS SISTEMAS (HOUGHTON ET AL, 2004)	28
FIGURA 5 - O LOOP OODA (HOUGHTON ET AL, 2004)	28
FIGURA 6 - OS DIFERENTES ENFOQUES PARA O BUSINESS INTELLIGENCE (ECKERSON, 2007)	31
FIGURA 7 - O CICLO DE VIDA DAS DECISÕES (HACKATHORN, 2004)	33
FIGURA 8 - O IMPACTO NA REDUÇÃO DAS LATÊNCIAS (HACKATHORN, 2004)	33
FIGURA 9 - A TOMADA DE DECISÃO EM TEMPO REAL (PANIAM, 2007)	36
FIGURA 10 - ESTÁGIOS NA ADOÇÃO DO REAL-TIME BUSINESS INTELLIGENCE (ECKERSON, 2007)	40
FIGURA 11 - EXEMPLO DE ARQUITETURA (GOLFARELLI; RIZZI; CELLA, 2004)	45
FIGURA 12 - ABORDAGEM CONVENCIONAL VERSUS IN-MEMORY (HOWSON, 2009)	56
FIGURA 13 - PROCESSAMENTO "INBOUND" VERSUS "OUTBOUND" (STONEBRAKER; ÇETINTEMEL,	
2005)	
FIGURA 14 - ARQUITETURA DA QUICKEN LOANS (ECKERSON, 2006)	
FIGURA 15 - ARQUITETURA DE INTEGRAÇÃO DE DADOS EM TEMPO REAL DO <i>YAHOO!</i> (AHUJA <i>ET AL</i> , 20	,
FIGURA 16 - OS ELEMENTOS QUE COMPÕE A LATÊNCIA NOS DADOS	
FIGURA 17 - O MODELO DE IDENTIFICAÇÃO DE MCGEE (MCGEE, 2004)	
FIGURA 18 - O MODELO DE JUSTIFICAÇÃO DE MCGEE (MCGEE, 2004)	
FIGURA 19 - ESTABELECENDO UMA VISÃO PARA O REAL-TIME BUSINESS INTELLIGENCE	
FIGURA 20 - Latência dos dados versus os custos e o valor para o negócio (ECKERSON, 20	,
FIGURA 21 - A IMPORTÂNCIA DE ALGUMAS TECNOLOGIAS NA ADOÇÃO DO BI OPERACIONAL, OU REAL-TI	
BUSINESS INTELLIGENCE (ECKERSON, 2007)	
FIGURA 22 - REDE DE ATMS	
FIGURA 23 - AMBIENTE ATUAL DE BUSINESS ACTIVITY MONITORING	
FIGURA 24 - PRIMEIRA ALTERNATIVA (USO DE STREAM PROCESSING E ODS)	
FIGURA 25 - SEGUNDA ALTERNATIVA (USO EXCLUSIVO DO STREAM PROCESSING, SEM O USO DO ODS)	
FIGURA 26 - PROVA DE CONCEITO	101

# **LISTA DE QUADROS**

QUADRO 1 - TERMOS ASSOCIADOS A REAL-TIME BUSINESS INTELLIGENCE	11
QUADRO 2 - NÍVEIS DE MONITORAÇÃO	17
QUADRO 3 - COMPARATIVO DAS DIFERENTES FORMAS DE CAPTURA DOS DADOS	51
QUADRO 4 - Tipos de Variação para os <i>Dashboards</i> (FEW, 2006)	62
QUADRO 5 - Os três tipos de <i>dashboards</i> (ECKERSON, 2006)	63
QUADRO 6 - EXEMPLO DE MODELO DE IDENTIFICAÇÃO E JUSTIFICAÇÃO PARA UMA EMPRESA DO SETOR DE	E
TRANSPORTE AÉREO.	73
QUADRO 7 - OBJETIVOS DE NEGÓCIO A PARTIR DA MATURIDADE DOS PROCESSOS (PANIAM, 2007)	74
QUADRO 8 - OPÇÕES TECNOLÓGICAS PARA A AQUISIÇÃO DE DADOS	83
QUADRO 9 - LATÊNCIA E RETENÇÃO NA PRIMEIRA ALTERNATIVA	99
QUADRO 10 - LATÊNCIA E RETENÇÃO NA SEGUNDA ALTERNATIVA	100
QUADRO 11 - COMPARATIVO ENTRE ALTERNATIVAS	. 100

# **LISTA DE TABELAS**

TABELA 1 - EXEMPLO DE PONTOS DE DADOS (BAM)	18
TABELA 2 - EXEMPLOS DE REGISTROS (BAM)	
TABELA 3 - CURRENT VALUE TABLE (BAM)	19
TABELA 4 - PERFIL DE ATUALIZAÇÃO NO DW EM 2003 (AGOSTA, 2003)	34
TABELA 5 - PERFIL DE ATUALIZAÇÃO DO DW EM 2007 (ECKERSON, 2007)	35

#### LISTA DE ABREVIATURAS E SIGLAS

API Application Program Interface ATM Automated Teller Machine BAM Business Activity Monitoring

ВΙ Business Intelligence

**BPM** Business Performance Management

CDC Change Data Capture

CEP Complex Event Processing

CVT Current Value Table

DEC Digital Equipment Corporation DSS Decision Support Systems

DW Data Warehouse

EAI Enterprise Application Integration **EDW** Enterprise Data Warehousing ΕII Enterprise Information Integration **EIS** Executive Information Systems **ERP** Enterprise Resouce Planning

**ETL** Extract-Transform-Load **GPS** Global Positioning System KPI Key Performance Indicator **MPP** Massive Parallel Processing ODS Operational Data Store

On-Line Analytical Processing OLAP **OLTP** On-Line Transaction Processing

OODA Observar, Orientar, Decidir e Agir

OPC **OLE for Process Control RAM** Random Access Memory RIA Rich Internet Application

RTBI Real-Time Business Intelligence

RTE Real-Time Enterprise RTI Real-Time Integrator

SQL Structured Query Language STP Straight Throught Processing ΤI Tecnologia da Informação **VIS** Vigilant Information Systems XML Extensible Markup Language

ZLE Zero-Latency Enterprise

# 1 INTRODUÇÃO

A demanda por maior previsibilidade no planejamento e execução dos negócios remonta à década de mil novecentos e oitenta. No âmbito acadêmico, algumas idéias motivaram a adoção de novas técnicas e modelos para a realização de negócios aplicando novos conceitos como planejamento integrado, redução de estoques e controle/automação de operações. Dentre essas idéias, pode-se citar o *Just-in-Time*, a re-engenharia de processos de negócios e o mais recentemente *Supply-Chain Management*.

Os novos modelos de gestão e a maior competição em alguns mercados contríbuíram para o aumento da complexidade no planejamento e na execução dos negócios. As empresas estão mais conectadas a seus fornecedores e clientes. O ambiente mais complexo introduz novas variáveis pouco trabalhadas.

#### Segundo McGee (2004a):

"diante desse novo cenário, as empresas precisam rever a forma de encarar algumas questões como: 1) a imprevisibilidade de certos eventos e o impacto que produzem nos negócios; 2) a inexistência de dados para a previsão de certos eventos; 3) existem muitos dados e isso dificulta a análise de cenários; 4) a incapacidade de prever certos eventos".

Em primeiro lugar, o acesso mais fácil à informação e as novas tecnologias de armazenamento e processamento permitem a construção de modelos de avaliação de cenários e riscos que permitem prever certos tipos de eventos com bastante precisão. Em segundo lugar, os dados sempre estiveram disponíveis e sempre estarão. As empresas precisam investir em mecanismos para capturá-los, armazená-los e tirar proveito deles. Em terceiro lugar, na prática, na maior parte dos casos conhecidos o que se observa é que o volume de dados necessários para o aumento de previsibilidade sobre certos eventos é relativamente pequeno comparado à infinidade de dados disponíveis. Em último lugar, adquirir a capacidade de prever certos eventos e tirar proveito deles na execução dos negócios passa a

ser uma nova competência que pode ser adquirida pelas empresas, desde que estas estejam dispostas a isso.

Adquirir a capacidade de capturar novos tipos de dados, armazená-los e distribuí-los de forma eficiente de modo que os gestores tirem proveito da informação vem sendo o desafio de todas as empresas que investem na construção ou no aprimoramento de um *Data Warehouse*. Nos últimos anos, esse desafio tornou-se mais complexo devido ao fato das empresas terem que lidar com diferentes e novos tipos de dados, estruturados ou não e terem que processar volumes maiores de dados provenientes de fontes diversas. Muitas empresas, devido ao seu ambiente de negócios altamente competitivos são obrigadas a fazer tudo isso em intervalos cada vez menores de tempo.

Nos últimos anos o *Data Warehouse*, como mecanismo de integração de dados mostrou-se insuficiente para atender às novas demandas do negócio. Novos mecanismos de integração de dados surgiram e a adoção de um conjunto de mecanismos, aliados ao *Data Warehouse* vem se mostrando a alternativa mais viável para suportar a inteligência empresarial em tempo real.

#### 1.1 Justificativa

Até o presente momento, constata-se a existência de poucas publicações científicas tratando de questões como a inteligência empresarial em tempo real e a importância que a integração de dados tem na sua adoção.

A partir desta necessidade, surgiu a idéia do desenvolvimento desta dissertação para apresentar as tendências em inteligência empresarial em tempo real e o papel fundamental que a integração de dados tem na adoção e no desenvolvimento de novos sistemas de apoio e automação de decisões.

Uma distinção importante que se deve fazer é que o termo tempo real utilizado nesta dissertação refere-se a realizar as tarefas computacionais em intervalos compreendidos na faixa de poucos segundos, sendo que a não realização dessas

tarefas computacionais não deve trazer consequências "trágicas" ao sistema. Essa distinção se faz necessária, pois o termo tempo real é utilizado com frequência em sistemas nos quais a não realização de uma atividade pode causar danos, como por exemplo a colisão de um trem ou ainda o superaquecimento do forno de uma siderúrgica.

## 1.2 Objetivos

O objetivo principal deste trabalho é propor a adoção de novas técnicas de integração de dados, aplicadas na captura, armazenamento e processamento em janelas de tempo bastante reduzidas, aqui denominadas de tempo real, para a realização da Inteligência Empresarial com enfoques táticos e operacionais, cuja demanda e aplicações em diversos ramos cresce ano a ano.

Esse objetivo pode ser alcançado tendo em vista a pesquisa dos principais mecanismos de integração de dados em tempo real que podem ser agregados ao *Data Warehouse* na construção de sistemas de apoio e automação de decisões, também em tempo real. A opção pela aplicação em um caso real se justifica pela oportunidade de por em prática a teoria estudada.

De forma a se atingir esse objetivo central, colocam-se outros objetivos complementares. O primeiro é realizar uma revisão da literatura em torno do próprio conceito de inteligência empresarial em tempo real, uma vez que, embora esse assunto possua ampla cobertura nas publicações não acadêmicas, ainda existem poucos trabalhos acadêmicos relacionados a esse assunto.

O segundo objetivo complementar é apresentar as tendências no ambiente de inteligência empresarial, em particular, para o enfoque mais operacional, para a necessidade de redução nas latências e nos níveis de maturidade possíveis de serem alcançados.

#### 1.3 Estrutura do Trabalho

Este capítulo contém uma introdução ao tema "inteligência empresarial em tempo real". Após isso, no capítulo 2, apresenta-se uma revisão da literatura em torno desse conceito buscando colocar as diferentes classificações usadas no meio não acadêmico e ao mesmo tempo estabelecendo ligações com trabalhos acadêmicos.

O capítulo 3 apresenta as principais tendências no ambiente de inteligência empresarial, dando ênfase às demandas táticas e operacionais. O capítulo 4 apresenta os principais mecanismos de integração de dados que, somados ao *Data Warehouse*, podem ser utilizados na construção de sistemas de apoio e automação de decisões. O capítulo 4 também apresenta dois casos reais de aplicação que visam ilustrar a aplicação dos conceitos sobre integração de dados apresentados nesse capítulo.

No capítulo 5, apresentam-se as principais considerações e decisões a serem levadas em conta na escolha dos mecanismos de integração de dados para suportar a inteligência empresarial em tempo real.

O capítulo 6 contém a descrição de um relato de caso implementado. Finalmente, o capítulo 7 contém as conclusões do trabalho.

## 2 A INTELIGÊNCIA EMPRESARIAL EM TEMPO REAL

A partir do final da década de mil novecentos e oitenta, começa a difundir-se no mercado de tecnologia da informação e nas empresas o conceito de *Business Intelligence* (ou Inteligência Empresarial). Esse novo conceito, seguido pelo novo acrônimo "BI", na forma como é conhecido nos dias de hoje, foi cunhado em 1989 por Howard Dresner, na época, executivo de marketing da *Digital Equipment Corporation* (DEC) e, posteriormente, analista do *Gartner Group* (WATSON; WIXOM, 2007). Dresner teve importante papel na disseminação do conceito durante seus anos no *Gartner Group*, embora o termo *Business Intelligence* já tivesse sido empregado por Hans P. Luhn (LUHN, 1958).

Na visão inicial de Dresner, tais sistemas englobam conceitos e métodos que permitem o aprimoramento do processo decisório através de fatos fornecidos por sistemas de suporte à decisão. Por esse e outros motivos, muitos autores ainda tratam do *Business Intelligence* como uma evolução dos sistemas de suporte à decisão (ou *Decision Support Systems*).

Mais recentemente, a partir do final da década de mil novecentos e noventa, o Business Intelligence passa a oferecer aos usuários de negócio um acesso mais fácil às informações. Um avanço importante frente aos Decision Support Systems (DSS) e aos Executive Information Systems (EIS) é que o Business Intelligence passa a prover o acesso à informação no modelo de auto-serviço (ou self-service), sem que os usuários dependam das áreas de sistemas.

Dois fatores foram importantes para a difusão do conceito de *Business Intelligence* nas empresas desde então:

1) A adoção do *Data Warehouse* nas grandes empresas permitia às áreas de negócio obter uma visão integrada dos dados, sem que, para isso, fosse necessário executar consultas no ambiente de produção. Com isso, o *Data Warehouse* transformou-se rapidamente em fonte principal de informações, uma vez que os usuários podiam executar suas consultas com maior

liberdade, sem que isso ocasionasse impacto nos demais sistemas. No início, como a maior parte dos *Data Warehouses* era construída sobre um banco de dados relacional, os usuários precisavam ter conhecimento sobre a linguagem *SQL* para terem acesso a ele. Posteriormente, esse acesso foi facilitado por sofisticadas ferramentas de análise e geração de relatórios;

2) Ao final da década de mil novecentos e noventa, com o desenvolvimento da computação pessoal, das redes de computadores e da adoção crescente do modelo cliente-servidor, surgem as primeiras ferramentas de análise e geração de relatórios. Elas foram chamadas de ferramentas OLAP (acrônimo criado por Edgar F. Codd em 1993 para On-Line Analytical Processing) e permitiram que um grupo maior de usuários fizesse uso dos dados no Data Warehouse, uma vez que, não era mais preciso ter domínio da linguagem SQL. Essa simplificação no acesso aos dados permitiu que mais usuários tirassem proveito do Data Warehouse. Consequentemente, a abrangência do Data Warehouse cresceu, deixando de atender apenas as áreas financeira e comercial para atender praticamente toda a empresa.

De acordo com o modelo proposto por Watson e Wixom (2007), o *Business Intelligence* pode ser caracterizado por um fluxo de retirada dos dados a partir do *Data Warehouse* sendo que este é caracterizado por um fluxo de entrada de dados a partir dos sistemas transacionais, conforme ilustra a figura 1.

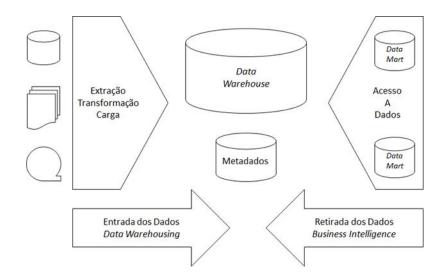


FIGURA 1 - Um modelo para o Business Intelligence (WATSON; WIXOM, 2007)

#### Segundo Loshin (2003):

"O *Business Intelligence* engloba o conjunto de processos, tecnologias e ferramentas necessárias para que as organizações transformem dados em informações, informações em conhecimento e conhecimento em planos que resultem em ganhos para o negócio. O *Business Intelligence* engloba o *Data Warehouse*, as ferramentas *OLAP*, as ferramentas de gestão de conteúdo e de conhecimento".

Nos últimos anos, o conceito de *Real-Time Business Intelligence* (ou inteligência empresarial em tempo real) ganhou atenção especial de revistas e jornais especializados em tecnologia da informação. Esse conceito muitas vezes é associado à idéia de empresa em tempo real (ou *Real-Time Enterprise*). Alguns autores também utilizam a expressão "empresa de latência zero" (ou *Zero Latency Enterprise*). Embora não se trate de uma idéia originalmente nova, muitos autores como (HACKATHORN, 2004), (MCGEE, 2004a), (REDDY, 2004), (SAWHNEY, 2003) e (WHITE, 2003) a apresentam como uma nova tendência num ambiente de negócios cada vez mais competitivo, no qual boa parte das empresas encontram-se inseridas.

#### Segundo McGee (2004a):

"a empresa em tempo real é capaz de monitorar, capturar e analisar eventos observáveis, de causa-raiz, que são críticos para o seu sucesso, imediatamente após a ocorrência dos mesmos, permitindo assim a identificação de novas oportunidades, evitando falhas e minimizando atrasos em processos chave. A empresa em tempo real deve explorar essas capacidades de modo a eliminar todos os atrasos na gestão e execução de seus processos críticos de negócio".

As idéias que motivam os negócios em tempo real surgiram há algumas décadas. Dentre elas pode-se citar a competição baseada no tempo (STALK, 1988), o *Just-in-Time* e a re-engenharia de processos de negócios (SAWHNEY, 2003). Mais recentemente, pode-se citar o *Total Quality Management* e o *Balanced Scorecard* (KAPLAN; NORTON, 1992).

#### Segundo Reddy (2004):

"o conceito de empresa em tempo real está presente há algum tempo. O objetivo inicial dos sistemas de informações era capturar dados transacionais e operacionais à medida que eram gerados, e ao mesmo tempo, compartilhá-los instantaneamente de modo a melhor suportar a cadeia de valor. As limitações tecnológicas fizeram a idéia de negócios em tempo real parecer uma ilusão. Embora muitas tecnologias de tempo real estivessem à disposição há algum tempo, a sua aplicação esteve sempre restrita a algumas funções do negócio como, por exemplo, a verificação de crédito nos pontos de venda. Mais recentemente, com uma maior disponibilidade das redes de alta velocidade e de novas tecnologias, é possível efetuar uma quantidade maior de funções do negócio em tempo real".

Embora a motivação para a adoção do *Real-Time Business Intelligence* exista há bastante tempo, até há pouco, haviam questões como a limitação tecnológica e o alto custo de adoção que impediam a sua realização. Muitas tecnologias que antes não existiam ou que apresentavam um alto custo, hoje estão à disposição das empresas a um custo mais acessível. Apesar do cenário mais favorável para investir em novas tecnologias, muitas empresas ainda enfrentam dificuldades na adoção do *Real-Time Business Intelligence*.

#### De acordo com McGee (2004b):

"por volta de cinco por cento dos processos de negócios de uma empresa são suficientemente críticos a ponto de exigirem a sua execução em tempo real. Adaptar os sistemas da empresa para operarem em tempo real pode exigir investimentos bastante altos. Ao mesmo tempo, se a empresa não for capaz de adquirir as competências de monitoração, análise e execução em tempo real, todo o investimento não renderá os benefícios esperados".

#### 2.1 Revisão da Literatura

Ao contrário do que ocorreu com *Data Warehouse*, onde a sua disseminação no meio empresarial foi posterior à sua formalização e aceitação no meio acadêmico, no caso do *Business Intelligence* e, consequentemente, do *Real-Time Business Intelligence* ocorreu justamente o contrário. Se por um lado hoje o *Business Intelligence* faz parte do dia-a-dia das grandes empresas, por outro ainda não existe uma definição deste conceito que seja aceita pela academia de forma ampla. Um dos fatores que possivelmente contribuiu neste processo foi o uso indiscriminado do acrônimo "BI" e da expressão *Business Intelligence* em contextos muito variados por parte dos meios de comunicação convencional, tais como os jornais e as revistas especializadas no mercado de tecnologia da informação. Para piorar a situação, alguns fornecedores de tecnologia buscaram adaptar o conceito em benefício próprio.

Com relação ao conceito de *Real-Time Business Intelligence* o que se observa nos meios convencionais é uma variação de acrônimos ou nomes que muitas vezes referem-se às mesmas coisas. Dentre as principais variações encontradas podem ser citadas: *Operational Business Intelligente*, *Near Real-Time Business Intelligence* e *Right-Time Business Intelligence*. Outras variações são: *Active Business Intelligence*, *Business Activity Monitoring*, *Business Performance Management* e *Proactive Business Intelligence*.

No meio acadêmico, são poucos autores que fazem uso do termo *Business Intelligence* e de suas variações. Em geral os autores preferem manter o uso do termo *Data Warehouse* e suas variações como: *Active Data Warehouse*, *Real-Time Data Warehouse* ou *Near Real-Time Data Warehouse*.

Ao mesmo tempo em que não existe uma norma para se tratar o assunto *Business Intelligence* no meio acadêmico, há pouquíssimos trabalhos buscando uma melhor estruturação do assunto. De toda a literatura considerada nesta dissertação, o trabalho mais completo no sentido de se estabelecer as bases de comparação entre as diferentes abordagens, considerando-se tanto a visão prática quanto a acadêmica, encontra-se no artigo de Greta L. Polites (POLITES, 2006). Um trabalho que busca estabelecer uma visão geral sobre o tema *Real-Time Business Intelligence* é o artigo de Azvine et al. (AZVINE *et al*, 2006). O outro trabalho que busca avaliar a evolução do *Business Intelligence*, a partir do *Data Warehouse*, é o trabalho de Golfarelli et al. (GOLFARELLI; RIZZI; CELLA, 2004).

Devido à ausência de uma concreta formalização no meio acadêmico com relação ao assunto *Real-Time Business Intelligence*, neste trabalho utilizou-se a busca de referências no meio acadêmico por trabalhos relacionados a sinônimos ou a temas mais específicos. Desta forma, partiu-se de jornais ou revistas sem credibilidade científica, buscando estabelecer ligações entre conceitos oriundos da esfera prática com trabalhos acadêmicos. A partir desta pesquisa, apresenta-se a seguir, uma compilação de temas relacionados ao assunto *Real-Time Business Intelligence*. Como forma de classificação, utilizou-se o quadro 1, que foi adaptado de (POLITES, 2006):

Termo	Descrição
Active Data Warehousing	O Data Warehouse é visto como uma ferramenta que deve ter uma participação ativa na realização dos negócios da empresa.  O Active Data Warehouse deve suportar a "Empresa Inteligente" que por sua vez responde mais rapidamente às demandas do mercado (HACKATHORN, 2002).
Business Activity Monitoring	Um processo a partir do qual eventos operacionais chave para o negócio são monitorados tal que, mudanças nos mesmos ou tendências, indiquem oportunidades ou problemas de modo que os gestores possam tomar as ações corretivas (NESAMONEY, 2004).
Business Performance  Management	Inclui todos os processos, metodologias, métricas e tecnologias que as empresas utilizam para medir, monitorar e gerenciar o desempenho do negócio (BUYTENDIJK; WOOD; GEISHECKER, 2004).  É a junção de planejamento, orçamento e KPI's utilizando-se um mesmo arcabouço estratégico e tecnológico que direcione todas as partes da empresa a atingir metas e objetivos (ECKERSON, 2004a).
Proactive Business Intelligence	Envolve "tempo real, missão crítica, planos de ação gerados automaticamente a partir de consultas ao <i>Data Warehouse</i> que são comunicados de forma proativa, utilizando-se a comunicação sem fio, de modo a fazer com que as pessoas certas tomem as ações corretas" (LANGSETH; VIVATRAT, 2002).
Vigilant Information Systems	Um sistema que é capaz de integrar e detalhar informações a partir de diversas origens de modo a detectar mudanças, enviar alertas, auxiliar no diagnóstico e análise de problemas para uma tomada de ação ágil. Os sistemas VIS devem incluir capacidades que o permitam sentir e responder. A capacidade de sentir é proporcionada por <i>dashboards</i> em tempo real e por alertas. A capacidade de responder é proporcionada através do apoio a decisão e planejamento de ações. (HOUGHTON <i>et al</i> , 2004)

QUADRO 1 - Termos associados a Real-Time Business Intelligence

A partir da compilação dos termos, a próxima atividade deste trabalho consiste em estabelecer um conjunto de características comuns às diferentes visões existentes

para o conceito de *Real-Time Business Intelligence* com o objetivo de estabelecer as bases para uma nova definição do assunto.

#### 2.1.1 Active Data Warehousing

#### Segundo Hackathorn (2002):

"A forma como o *Data Warehouse* é empregado na condução dos negócios da empresa pode determinar significativamente o seu sucesso. O uso ativo desse recurso como o motor das ações ao invés de simples fonte de informação é a chave por trás do *Active Data Warehousing*".

#### Segundo Whitehorn (2008):

"O Active Data Warehousing nasce a partir das necessidades do negócio ou das áreas operacionais da empresa e não como um modismo ou tendência do mercado".

Apesar do termo *Active Data Warehousing* ter sido empregado por alguns fornecedores como a Teradata, alguns analistas de mercado o consideram como um processo e não uma tecnologia (AGOSTA, 2004).

Dentre as condições necessárias para que um *Data Warehouse* possa ser classificado como *Active Data Warehouse* estão:

Visão singular da verdade (ou Single Version of True). A partir desta idéia o Data Warehouse deve conter todos os dados necessários à execução dos negócios. Esses dados devem ser consistentes com a realidade e acurados. O Data Warehouse deve ser a principal fonte de informações para a análise e tomada de decisão. Em muitas empresas, apesar da utilização do Data Warehouse, algumas informações encontram-se dispersas em outras fontes de dados. Se para realizar alguma análise relevante ao negócio for necessário cruzar dados do Data Warehouse com o de algum Data Mart

então esse *Data Warehouse* não pode ser considerado um *Active Data Warehousing* (AGOSTA, 2004; HACKATHORN, 2002);

- Granulosidade dos dados adequada, ou seja, que permita obter o detalhamento necessário dos dados na hora de tomar as ações (HACKATHORN, 2002).
- Perfil de consultas misto (ou Mixed Workloads). No Active Data Warehousing deve haver tanto as consultas operacionais que objetivam o processo de decisão tático quando as consultas extremamente complexas que geralmente buscam atender às demandas mais estratégicas e de longo prazo. A concentração ou predominância de um dos perfis, operacional ou estratégico, é um indicador de que não se trata de um Active Data Warehouse (AGOSTA, 2004);
- Tempestividade na atualização dos dados a partir dos sistemas de origem (HACKATHORN, 2002; PUTTANGUNTA, 2006);
- O *Active Data Warehouse* deve gerar alertas e ações e transmiti-las automaticamente aos sistemas operacionais da empresa (AGOSTA, 2004).

Diante dessa visão, o *Active Data Warehouse* passa a ter um papel de destaque no processo decisório e na execução das atividades críticas do negócio.

No âmbito acadêmico, praticamente não existe menção ao termo *Active Data Warehousing; entretanto*, existem muitos trabalhos relacionados aos problemas e desafios para a implantação de um *Active Data Warehousing*, tais como:

- Manutenção de maior granulosidade dos dados (BRUCKNER; LIST; SCHIEFER, 2002; CONN, 2005);
- Perfil misto de consultas (TAYLOR, 2000; BRUCKNER; LIST; SCHIEFER, 2002; CONN, 2005; STONEBRACKER; ÇETINTEMEL, 2005; MATIAS, 2006; DAYAL et al, 2009);

 Atualização dos dados em tempo real (BRUCKNER; LIST; SCHIEFER, 2002; KARAKASIDIS; VASSILIADIS; PITOURA, 2005; STONEBRACKER; ÇETINTEMEL, 2005; ITALIANO; FERREIRA, 2006; SANTOS; BERNARDINO, 2008; CASTELLANOS et al, 2009).

#### Segundo Bruckner et al. (2002):

"Transformar um ambiente de *Data Warehouse* de modo a suportar a atualização dos dados em tempo real exige a revisão de níveis de serviço tal que o mesmo exceda os níveis comumente acordados".

Para que se alcance a atualização dos dados em tempo real é preciso:

- Realizar a integração dos dados de forma contínua através da captura de atualizações e carga de dados a partir dos sistemas de origem em tempo real;
- Ter à disposição mecanismos de tomada de decisão que trabalhem com regras ou com base em eventos de modo a automatizar parte do processo decisório;
- Ter à disposição um ambiente que permita a análise de grandes volumes de dados de forma ininterrupta, ou seja, sem a existência de janelas para cargas de dados.

# 2.1.2 Business Activity Monitoring

A análise e a visualização de dados em tempo real não é uma necessidade recente. Ela surgiu com a adoção dos sistemas de controle de processos e automação em manufatura e sistemas de telemetria. Tais sistemas existem há algumas décadas. Mais recentemente, observou-se um maior interesse na monitoração e controle de aplicações empresariais, analogamente ao que se faz nos sistemas de controle de processos.

As aplicações que permitem a monitoração de atividades, a avaliação do status e a saúde de sistemas empresariais denominam-se *Business Activity Monitoring* ou simplesmente BAM, como acrônimo. Na prática, o objetivo desses sistemas é advertir sobre a ocorrência de determinados eventos relevantes para o negócio. De forma análoga ao que se realiza no controle de processos de manufatura, o que se espera com a adoção das ferramentas de BAM é realizar algo similar com os principais processos da empresa. Tais sistemas exigem a análise dos dados em tempo real, ou seja, mais uma característica comum aos sistemas de controle de processo. Tanto nos sistemas de controle de processo quanto nos sistemas BAM existe a necessidade de aquisição de dados em tempo real, a necessidade de análise dos dados e a apresentação dos dados em formas variadas.

#### De acordo com o Gartner Group:

"Um sistema de *Business Activity Monitoring* oferece acesso em tempo real a indicadores de desempenho críticos ao negócio com o objetivo de acelerar as operações de negócio e ao mesmo tempo torná-las mais eficientes" (MCCOY, 2002).

"O *Business Activity Monitoring* não se trata de uma evolução do *Business Intelligence* ou de um substituto e sim de uma complementação" (DRESDNER, 2002).

O papel do *Business Activity Monitoring* é o de reduzir ou eliminar atrasos, gargalos e o uso ineficaz de mão de obra e materiais enquanto provê dados de desempenho dos processos em tempo real. O BAM se transformará na ferramenta analítica a ser empregada pelas empresas interessadas em trabalhar com latência zero, ou seja, reduzindo o tempo entre a captura de eventos críticos ao negócio e a sua notificação para que ações sejam tomadas (GOVEKAR; SCHULTE, 2002).

#### De acordo com Luckham (2004):

"Um sistema de Business Activity Monitoring deve ser capaz de:

- Advertir instantaneamente sobre eventos nos sistemas de informações em qualquer nível (ex. falhas na rede, atividades no web-site, alteração de valores em bases de dados, etc). Devem ser considerados quaisquer eventos que afetem a execução dos processos de negócios críticos;
- Permitir a tomada de decisão em tempo real em resposta aos eventos capturados a partir dos sistemas de informação;
- Automatizar as notificações de violação de políticas;
- Prover estatísticas sobre o desempenho do negócio".

Um sistema de *Business Activity Monitoring*, ao contrário de outras ferramentas de monitoração disponíveis em uma empresa, concentra-se em fontes de dados que permitem obter os indicadores chave de negócio (ou KPI's), embora também possa se servir de dados de outras fontes. O quadro 2 apresenta uma comparação do BAM com outras funções de monitoramento que possam existir no ambiente de tecnologia da informação tais como a monitoração de serviços e a monitoração da infraestrutura.

Nível	Fontes de Dados	Indicadores de Desempenho		
Business	Pacotes ERP, aplicações	Quantidade de Pedidos no dia,		
Activity Monitoring	desenvolvidas	Quantidade de Solicitações dos		
	internamente, fontes de	Clientes, Quantidade de		
	informações externas,	Reclamações		
	sistemas de mensageria,			
	bancos de dados, agentes			
	de monitoração, redes de			
	sensores.			
Monitoração de	Pacotes ERP,	Mensagens		
Serviços	Sistemas de Mensageria,	enviadas/recebidas,		
	Banco de Dados	Quantidade de Acessos		
Monitoração de	Servidores, Agentes de	Consumo de CPU, de		
Infraestrutura	Monitoração	memória, etc.		

QUADRO 2 - Níveis de Monitoração

Por muito tempo a monitoração de dados em tempo real esteve associada, geralmente, a aplicações como a telemetria e ao controle automático de processos industriais. Nesse tipo de aplicação, os dados eram gerados a partir de sensores (ex. temperatura ou pressão) e transmitidos a um servidor central em intervalos regulares de tempo, utilizando-se para isso algum protocolo padrão como, por exemplo, o OPC (ou *OLE for Process Control*).

Ao contrário dessas aplicações, onde existem soluções bastante avançadas, quando se trata da monitoração de aplicações empresariais, a situação é um pouco diferente. Em primeiro lugar, falta uma padronização com relação às fontes de dados, uma vez que, as aplicações existentes, geralmente, foram desenvolvidas em momentos diferentes e utilizando tecnologias e metodologias distintas. A falta de padrões e a falta de integração entre tais sistemas, por si só, já representa um grande desafio. Neste contexto, surgiram as ferramentas de integração de aplicação, também conhecidas como *Enterprise Application Integration*. Tais ferramentas, por sua vez, concentram seus esforços na resolução do problema da falta de integração

entre as aplicações, porém oferecem pouco suporte à monitoração dos eventos de negócios em tempo real.

Assim como a captura dos dados difere entre os sistemas de controle de processos e os sistemas BAM, o próprio formato dos dados também difere. Os sistemas de controle de processos operam com sequências de pontos de dados onde cada ponto possui uma marca no tempo (ou *timestamp*) e um valor associado a uma determinada medida (ex. temperatura, conforme a tabela 1).

Timestamp	Temperatura
7:00	35,2
7:01	35,4
7:02	35,7
7:03	36,1
7:04	36,6

TABELA 1 - Exemplo de Pontos de Dados (BAM)

Nos sistemas BAM, devido às fontes de dados dos sistemas empresariais geralmente trabalharem com bancos de dados relacionais e, em função do uso das plataformas de integração de aplicações (EAI), os dados são capturados de forma mais estruturada, comumente sob a forma de registros em uma tabela, tais como ilustrado na tabela 2:

Timestamp	Região	Funcionário	Operações	Cancelamentos	Falhas
7:05	Norte	João	3	0	1
7:05	Sul	Maria	4	1	0
7:10	Norte	João	4	1	0
7:10	Sul	Maria	4	0	1
7:10	Oeste	Luis	3	0	0

TABELA 2 - Exemplos de Registros (BAM)

Um aspecto importante dos sistemas de *Business Activity Monitoring* é a necessidade de persistência dos dados quando comparados a seus precursores (sistemas de controle e telemetria). Os dados capturados devem ser utilizados em tempo real mas também devem ser armazenados por um determinado período para que possam ser empregados em análises.

Outro aspecto importante dos dados é que os eventos em tempo real são capturados de forma assíncrona, ou seja, podem ser capturados fora de sua ordem cronológica a partir das diversas fontes. Essa característica produz um problema. Por exemplo, o escritório central da empresa pode receber mensagens com as transações realizadas em cada filial em instantes diferentes. Cada mensagem origina um registro contendo muitas dimensões das medidas avaliadas. Assim sendo, fica complicado determinar, a partir das mensagens, o que está acontecendo.

A solução mais utilizada na resolução desse problema é a manutenção de uma tabela com os valores correntes (ou *Current Value Table*, ou CVT) onde as colunas são indexadas de modo a identificar de forma única a origem dos dados. Os dados desta tabela são atualizados sempre que uma nova mensagem é recebida de modo que ela sempre apresente os dados atualizados. Na tabela 3 há um exemplo de CVT para as operações descritas na tabela 2:

Região	Funcionário	Operações	Cancelamentos	Falhas
Norte	João	7	1	1
Sul	Maria	8	1	1
Oeste	Luis	3	0	0

TABELA 3 - Current Value Table (BAM)

Na maior parte dos sistemas de *Business Activity Monitoring* os dados capturados são armazenados em bases de dados relacionais para fins de análise; entretanto, é difícil manter uma tabela com os valores correntes (ou CVT) em uma base relacional. Neste caso, seria necessária a realização frequente de consultas contra conjuntos grandes de dados históricos de modo a determinar os valores correntes.

Isto provocaria uma sobrecarga no servidor de banco de dados e o tempo de resposta para se obter a tabela atualizada seria aumentado, o que inviabiliza esse tipo de abordagem. Na prática, o que se faz é manter uma tabela de valores correntes em memória.

Existem muitos requisitos para a construção de uma tabela de valores correntes o que pode tornar sua implementação bastante complexa. Dentre alguns requisitos vale citar a necessidade de eliminar registros antigos ou vencidos, além de se trabalhar com muitas dimensões e, consequentemente, grandes volumes de dados (LUBINSKY, 2008).

Na análise dos dados, utilizam-se as funções analíticas mais simples (ex: a soma, a média e o desvio padrão). Além disso, utilizam-se frequentemente as funções de decomposição (ex: "group by").

À medida que a necessidade de integração de dados em tempo real se faz mais presente e, com o aumento no volume de dados e da velocidade de atualização dos mesmos, alguns autores como Luckham (2004), Küng et al (2005) e Greiner et al (2006) sugerem que as plataformas CEP (ou Complex Event Processing) são mais adequadas para a implantação de soluções de Business Activity Monitoring do que as plataformas de EAI (ou Enterprise Application Integration).

# 2.1.3 Business Performance Management

### Segundo Axson (2007):

"O Business Performance Management engloba todos os processos, informações e sistemas utilizados pelos gestores na definição da estratégia, desenvolvimento de planos, acompanhamento da execução de atividades, previsão e apresentação de resultados, de modo que a organização alcance o sucesso sustentável, independentemente do que a definição de sucesso venha a ser".

#### Segundo Eckerson (2006):

"Um dos princípios do *Business Perforformance Management* é que as empresas precisam identificar as atividades que mais contribuem para o seu sucesso e, assim, executá-las com excelência. Em outras palavras, o seu propósito é auxiliá-las a se tornarem mais focadas, alinhadas e efetivas".

#### O The Data Warehouse Institute define o Business Performance Management como:

"Uma série de processos e ferramentas que viabilizam a execução de uma estratégia de negócios. Ele promove uma boa gestão ao permitir que os executivos, em todos os níveis da organização, compreendam, comuniquem e monitorem os fatores chave para o negócio. Como forma de complementar esta definição, o instituto define que o *Business Performance Management* não é :

- Uma tecnologia ou solução de software;
- Uma nova categoria de ferramenta de Business Intelligence;
- Uma ferramenta de planejamento, orçamento ou consolidação;
- Um conjunto de Dashboards ou Scorecards;
- Uma ferramenta de modelagem ou previsão;
- Um conjunto de KPIs (Key Performance Indicators)". (ECKERSON, 2004)

Apesar de não ser nenhum dos elementos listados anteriormente, a construção de uma solução de *Business Performance Management*, segundo o *The Data Warehouse Institute*, envolve a integração de todos eles.

O Business Performance Management, enquanto processo, é composto de um ciclo com quatro etapas, sendo responsável por transformar a estratégia da organização em ações que resultem em um desempenho superior (ECKERSON, 2006). As etapas, conforme ilustra a figura 2, são: 1) a estratégia; 2) o planejamento; 3) o monitoramento e análise e por final 4) a ação e ajuste.

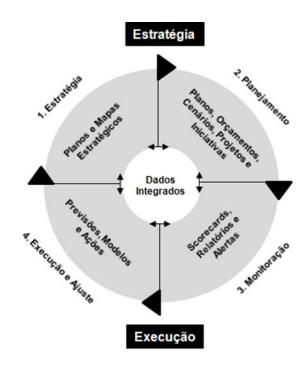


FIGURA 2 - O ciclo do BPM (ECKERSON, 2006)

Na primeira etapa, denominada estratégia, os executivos devem definir os fatores chave de sucesso (ou *key drivers*) e a forma como eles serão medidos. Um exemplo bastante comum de fator chave de sucesso é a satisfação dos clientes. Outro fator chave de sucesso que se pode citar é a qualidade dos produtos, que por sua vez pode ser medida, por exemplo, através da quantidade de defeitos em lotes de um mil. A definição dos fatores chave de sucesso não é tarefa exclusiva dos executivos; na verdade, todos os gestores responsáveis por unidades de negócios ou por departamentos devem participar. Em muitos casos, a avaliação de uma medida é influenciada por resultados isolados, podendo chegar até ao nível de cada indivíduo. Por exemplo, o índice de satisfação dos clientes pode ser avaliado por departamento, por unidade de negócio e depois consolidado como um único índice geral para toda a organização.

Denominam-se *KPI*s, ou *Key Performance Indicators* as medidas que permitem aferir quão bem uma unidade de negócios ou indivíduo está desempenhando. Os *KPI*s são identificados e definidos durante o planejamento estratégico. O acompanhamento e a manutenção desses indicadores dentro de limites desejáveis devem conduzir a organização a um desempenho superior.

Na segunda etapa, denominada de planejamento, os gestores desenvolvem os planos que determinarão a alocação dos recursos, de modo que os objetivos definidos anteriormente sejam cumpridos. Nessa fase, utilizam-se ferramentas de gestão, tais como: planejamento orçamentário, previsão de vendas e planos de projetos para avaliar e definir a melhor forma de alocar os recursos.

Em muitas organizações o planejamento ainda constitui uma atividade muito trabalhosa, onde os envolvidos comumente tomam decisões com base em fotografias do desempenho passado da empresa. Por se tratar de um processo de colaboração, o planejamento, atualmente, pode contar com a tecnologia como aliada; assim sendo, muitas empresas estão tirando proveito disso. Nestes lugares, as ferramentas de *Business Intelligence* e o *Data Warehouse* têm permitido a realização do planejamento de forma dinâmica, onde todos os envolvidos, nos diversos níveis hierárquicos, contam com recursos avançados que possibilitam a análise de diferentes cenários, permitindo visualizar o comportamento passado assim como a realização de projeções.

É importante frisar que, embora alguns autores relacionem o *Business Performance Management* à metodologia de *Balanced Scorecard*, não existe uma correspondência direta entre as duas coisas. Durante a execução das etapas de estratégia e planejamento podem ser utilizadas uma ou mais disciplinas existentes e relacionadas, tais como o Planejamento Estratégico Tradicional, o *Balanced Scorecard*, o *Six-Sigma*, o *Total Quality Management* e o *Economic Value Added* entre outros.

Na terceira etapa, denominada de Monitoração e Análise, a organização precisa se preparar para garantir a execução dos planos traçados. A avaliação dos *KPI*s deve ser constante, não se limitando apenas aos tradicionais relatórios de acompanhamento. Mais recentemente, com a utilização das novas ferramentas de *Business Intelligence*, também é possível contar com *dashboards*, *scorecards*, alertas e *cockpits*.

A quarta e última etapa é a mais crítica pois, para que a estratégia seja de fato executada, os colaboradores devem tomar decisões e ações de modo que qualquer

desvio frente ao planejado seja corrigido. É neste momento que o ferramental de Monitoração e Análise entra em ação. Esses mecanismos serão os responsáveis por alertar os colaboradores da ocorrência de problemas em potencial, além de oferecer informações atualizadas que apoiem as decisões táticas.

Dessa forma, encerra-se o ciclo. A utilização de dados integrados ao longo de todas as etapas é fundamental, uma vez que, possibilitará a geração de conhecimento e permitirá o seu aprimoramento, iteração após iteração. Também será visto, nesse caso, como se estabelecerá a dependência do processo de *Business Performance Management* com relação ao *Data Warehouse* e as demais soluções e arquiteturas para integração de dados.

O Business Performance Management implica em uma nova abordagem para o Business Intelligence. O fator principal que caracteriza essa mudança é a necessidade de uma capacidade de monitoração mais sofisticada, de modo que os processos de negócio mais críticos possam ser avaliados continuamente, permitindo que as decisões táticas e operacionais sejam embasadas em dados e fatos e não mais no simples julgamento de seus atores. A adoção do Business Performance Management como disciplina de gestão introduz novos desafios ao ambiente de Business Intelligence que vão desde a natureza dos dados, a sua coleta, a forma como a informação é extraída, assim como o seu ciclo de vida e tempestividade (GOLFARELLI; RIZZI; CELLA, 2004).

Um aspecto importante na abordagem do *Business Performance Management* é entender como as decisões são tomadas. Ao contrário do que ocorre comumente nas empresas, aquelas que adotam essa disciplina tendem a tratar o processo de decisão como um processo contendo um ciclo, como ilustra a figura 3:

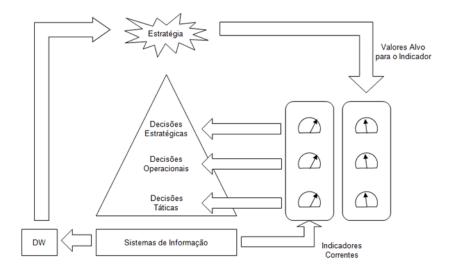


FIGURA 3 - O processo decisório (GOLFARELLI; RIZZI; CELLA, 2004)

Nessa abordagem, durante as fases de estratégia e planejamento, conforme apresentado anteriormente, são definidos os *KPI*s, ou indicadores objetivos e mensuráveis, cujo único propósito é avaliar o desempenho dos processos da organização. A ideia disso é traduzir a estratégia em um conjunto de indicadores, os quais, garantindo que se mantenham dentro de intervalos esperados, a organização produzirá um desempenho superior. Um aspecto importante é que os diversos indicadores, independentemente do processo ou nível hierárquico envolvido, estejam relacionados entre si de modo que todas as atividades contribuam para o resultado desejado.

Nesse modelo, a estratégia é elaborada com base nos dados e fatos obtidos a partir do *Data Warehouse*, além de outras fontes úteis enquanto a execução é realizada, levando-se em conta a situação corrente dos indicadores, ou seja, se os indicadores não estiverem de acordo com o planejado, então existe um problema e ações corretivas devem ser tomadas rapidamente.

Pode-se constatar que o elemento principal do *Business Performance Management* são os indicadores (ou *KPI*s). Algumas características importantes, listadas abaixo, distinguem uma abordagem de *Business Intelligence* voltada para o *Business Performance Management* de uma abordagem tradicional (GOLFARELLI; RIZZI; CELLA, 2004):

- Usuários: Neste caso, os usuários também são tomadores de decisão, mas atuam em um nível organizacional diferente daquele dos usuários tradicionais do *Data Warehouse*. Esses usuários possuem uma visão limitada da estratégia da empresa e, comumente, trabalham com conjuntos de indicadores relacionados a tarefas do seu cotidiano. O que se espera é que tais usuários sejam capazes de desempenhar suas atividades individuais de modo a contribuir para que os indicadores de desempenho relacionados a eles estejam sempre em conformidade. Dessa forma, estarão contribuindo para o sucesso da organização. Neste sentido, pode-se observar um volume significativamente maior de usuários quando comparado ao *Data Warehouse* tradicional, uma vez que, neste caso, pode-se falar de todos os colaboradores, do nível tático ao operacional.
- Latência dos Dados: No nível tático, as decisões devem ser tomadas com maior agilidade quando comparadas àquelas tomadas no nível operacional e estratégico. Desta forma, a latência nos dados deve ser adequada. Isso não significa que todos os dados devam ser atualizados instantaneamente (ou em tempo real). Ao contrário disso, pode-se dizer que os dados precisam ser atualizados no tempo certo (ou right-time). Esse intervalo de tempo pode variar em função de fatores como o negócio da empresa, o departamento em questão e o perfil do cliente. Ao contrário do que acontece no ambiente tradicional de Data Warehouse, as atualizações dos dados consumidos para a apresentação dos indicadores precisam ocorrer de forma contínua.
- Granulosidade e Ciclo de Vida dos Dados: Os dados utilizados na apresentação dos indicadores possuem uma granulosidade maior do que aquela tradicionalmente utilizada no Data Warehouse. Ao contrário do que acontece no ambiente de Data Warehouse, no qual as visões sumarizadas dos dados são suficientes, neste caso pode ser necessária a apresentação de dados detalhados em conjunto com as sumarizações realizadas nos últimos instantes. Outra característica importante é que, nesse caso, o que interessa é o estado atual das coisas e, portanto, não existe a necessidade de retenção dos dados por um período longo de tempo, ao contrário do que ocorre no Data Warehouse.

• Interface Gráfica e Perfil de Acesso aos Dados: Dashboards e alertas constituem a principal interface com os usuários. Em geral, os usuários não têm tempo ou habilidades suficientes para a utilização de ferramentas OLAP sofisticadas. Neste caso, a informação precisa ser entregue, na maioria das vezes, sob a forma de dashboards ou cockpits, onde a visualização dos indicadores possa ser realizada de forma simples e ágil.

### 2.1.4 Vigilant Information Systems

De acordo com o trabalho de Walls, Widmeyer e El Sawy (1992):

"O conceito de *Vigilant Information Systems* surge no início da década de mil novecentos e noventa, a partir de uma tentativa formal em se estabelecer uma teoria para os *Executive Information Systems* (ou EIS). O termo vigilância denota a habilidade de um sistema de informação em permitir ao corpo gerencial da empresa ficar em estado de alerta de modo a perceber os menores sinais e descontinuidades relevantes no ambiente de negócios, sejam oportunidades ou ameaças, do ponto de vista estratégico".

#### Segundo Houghton et al (2004):

"Ser vigilante, para um sistema de informação, significa estar em constante alerta o que inclui as capacidades de sentir e responder a estímulos. A capacidade de sentir está relacionada à detecção de mudanças no ambiente que permitam aos gestores obter uma melhor visão desde o chão de fábrica até os mais altos níveis da organização. Na prática, isto é realizado a partir do uso de painéis de informação (ou *dashboards*) em tempo real associados a mecanismos de alerta (visual, sonoro ou por mensagem). A capacidade de responder está relacionada ao apoio que tais sistemas oferecem aos gestores permitindo que tenham as informações necessárias no momento certo de tomarem decisões e ações".

Uma diferença importante para os *Vigilant Information Systems* é que não é o usuário quem inicia o processo de análise de informação como acontece nos sistemas tradicionais de suporte à decisão. Assim, como ilustra a figura 4, num *Vigilant Information System*, a partir de uma alteração no ambiente de negócios,

associada à capacidade de captura de dados que caracterizam essa mudança, é que se inicia o processo de análise.



FIGURA 4 - Vigilant Information Systems e os demais sistemas (HOUGHTON et al, 2004).

Um dos princípios por trás dos *Vigilant Information Systems* é o *loop* OODA que é um acrônimo para Observar, Orientar, Decidir e Agir (figura 5). O conceito do *loop* OODA foi desenvolvido na força-aérea americana, na década de mil novecentos e oitenta, quando o assunto em questão era o desempenho de diferentes pilotos e aeronaves. A idéia foi adaptada, posteriormente, para o ambiente de negócios.

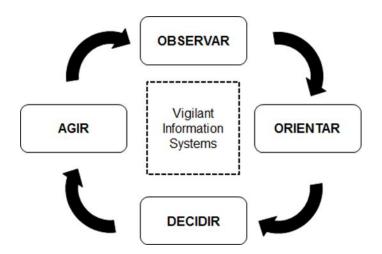


FIGURA 5 - O loop OODA (HOUGHTON et al, 2004).

Os princípios por trás do *loop* OODA se traduzem em capacidades a serem suportadas pelo sistema, tais como:

- Observação: O sistema deve prover visibilidade nos processos de negócio mais críticos da empresa. Além disso, deve ser capaz de capturar e atualizar os principais indicadores de desempenho (ou KPI's) em tempo real integrando dados das mais variadas fontes de dados;
- Orientação: O sistema deve fornecer painéis gráficos (ou dashboards) para a apresentação dos dados. Alertas devem ser enviados às pessoas responsáveis e a interface gráfica deve permitir analisar os dados sob diferentes ângulos e pontos de vista. Outros recursos importantes para facilitar a orientação dos usuários é o uso de indicadores coloridos (ou Trafficlight alerts) e relatórios apresentando curvas de tendência;
- Decisão: O sistema deve oferecer suporte analítico para que possam ser realizadas análises do tipo "what-if";
- Ação: O sistema deve permitir que as decisões sejam comunicadas de forma efetiva aos atores responsáveis por sua execução. Nos casos possíveis, o sistema deve automatizar a execução de ações através de integração com os demais sistemas da empresa.

## 3 TENDÊNCIAS EM INTELIGÊNCIA EMPRESARIAL

Os cenários de utilização do *Business Intelligence* vêm mudando nos últimos anos. Por um lado, uma maior regulamentação (ex: Lei Sarbarnes-Oxley em 2002) forçou muitas organizações a documentar melhor seus processos de negócio e a garantir a legitimidade das informações repassadas aos acionistas. Além disso, o ciclo de negócios mais apertado fez com que a tomada de decisões com mais qualidade e de forma mais rápida se transformasse em uma obrigação competitiva (TURBAN *et al*, 2008).

Com relação às tendências observadas para o *Business Intelligence* nos últimos anos, vale a pena destacar:

- Enfoque Operacional em crescimento;
- Necessidade por redução das latências;
- Tomada de Decisão em tempo real;
- Novas fontes de dados, internas e externas.

### 3.1 Enfoque Operacional

No início, os enfoques estratégico e tático predominaram na maior parte das organizações que adotaram o *Business Intelligence*. Como exemplos de decisões estratégicas destacam-se a introdução de um novo produto no mercado, a aquisição de um competidor, alterações nas estratégias de preços ou na força de vendas. Com relação às decisões táticas, pode-se citar a alocação de recursos para um novo orçamento, como alocar as verbas de promoção de modo a aumentar as vendas de um determinado produto ou como avaliar o impacto de uma alteração de preços. Com um enfoque mais operacional, o *Business Intelligence* começa a servir uma quantidade maior de colaboradores e de perfis funcionais, variando desde atendentes de *call center*, passando por técnicos em campo e chegando até a

parceiros de negócios. A seguir, destacam-se as principais características do enfoque operacional:

- O volume de decisões operacionais é muito maior do que as decisões estratégicas ou táticas;
- O valor individual das decisões operacionais é bem inferior ao valor individual das decisões táticas ou estratégicas; no entanto, a parcela de decisões operacionais comumente é representativa (ver figura 6);
- Os usuários precisam tomar suas decisões rapidamente (quando comparado às decisões táticas ou estratégicas). Em outras palavras, as informações necessárias para a tomada de decisão devem ser entregues no tempo e formato adequados, assim que um evento relevante acontece, para que os usuários do sistema tenham tempo hábil para reagir e tomar as ações adequadas;
- Em alguns casos, devido ao curto espaço de tempo disponível ou como forma de se aumentar a produtividade, algumas decisões operacionais podem ser realizadas de forma automática, sem a intervenção de usuários.

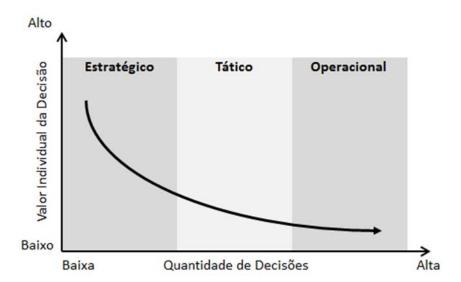


FIGURA 6 - Os diferentes enfoques para o Business Intelligence (ECKERSON, 2007).

### 3.2 Redução das Latências

Pode-se definir latência, de uma perspectiva do negócio, como o tempo que decorre desde o instante em que algum evento relevante ocorre até o momento em que alguma ação seja tomada, em consequência desse evento.

De acordo com modelo proposto por Hackathorn (2004), a latência pode ser classificada em três categorias ou tipos:

- Latência dos Dados;
- Latência da Análise;
- Latência da Decisão (ou da Ação).

A latência nos dados representa o tempo necessário para capturar, limpar, conformar, transformar e armazenar no *Data Warehouse* os dados relativos a uma transação de negócio de modo que esta fique disponível para a análise. A latência da análise engloba o tempo necessário para que os dados sejam analisados, empacotados em resultados (relatórios, indicadores, painéis) para que possam ser entregues às pessoas certas, ou seja, aquelas responsáveis pela tomada de decisão e ação. Por final, a latência da decisão (ou da ação) representa o tempo necessário para que o agente responsável pela decisão entenda a situação com base nos resultados da análise e tome a decisão (com posterior ação).

Cada um dos tipos de latência, por sua vez, implica em algum impacto ou resultado, do ponto de vista de valor para o negócio, conforme ilustra a figura 7.

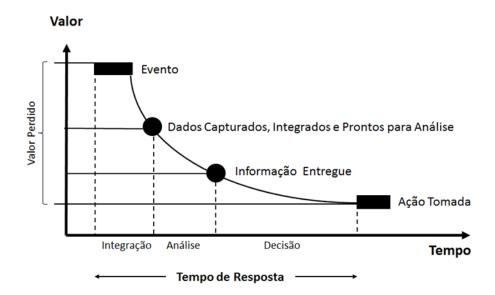


FIGURA 7 - O ciclo de vida das decisões (HACKATHORN, 2004)

Na maior parte das implantações de *Business Intelligence*, o que se procura fazer é reduzir a latência nos dados, seja através do tratamento de novos eventos de negócio não tratados pelo *Data Warehouse*, seja através de processos de atualização mais frequentes. Apesar dos esforços estarem mais direcionados à redução da latência nos dados, deve-se buscar reduzir todo o ciclo de modo a evitar a perda de valor para o negócio, assim como ilustra a figura 8.

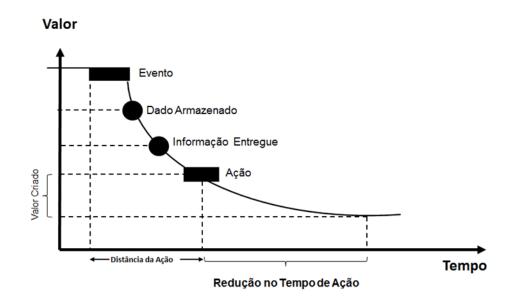


FIGURA 8 - O impacto na redução das latências (HACKATHORN, 2004)

Uma vez que todas as latências são reduzidas, existe potencial para a recuperação de valor para o negócio. O aspecto fundamental do modelo proposto por Hackathorn é que as ações empreendidas na redução das latências devem se basear no impacto positivo para o negócio que essas ações permitem.

A tendência pela redução das latências, sobretudo nos dados, pode ser observada por meio das pesquisas realizadas pelo *The Data Warehouse Institute*. Na primeira pesquisa, realizada no ano de 2003, observou-se uma predominância por processos de atualização diária no *Data Warehouse*, conforme apresenta a tabela 4:

Frequência de Atualização no DW	Ano de 2003	Previsão (18 meses seguintes)
Mensal	41%	27%
Semanal	26%	29%
Diária	75%	72%
Muitas vezes ao dia	2%	14%
Em tempo real	0%	10%

TABELA 4 - Perfil de Atualização no DW em 2003 (AGOSTA, 2003)

Como se observa na tabela 4, até 2003, houve uma predominância pela atualização diária no *Data Warehouse* com 75% das empresas realizando atualização diária e apenas 2% com atualização muitas vezes ao dia, ou intra-diária. Mesmo nos dezoito meses seguintes, segundo os entrevistados, ainda predominou a atualização diária, porém, com uma expectativa de crescimento das atualizações intra-diárias.

A partir de uma pesquisa similar, realizada pelo mesmo instituto em 2007, é possível observar o crescimento nas atualizações intra-diárias, como apresenta a tabela 5:

Frequência de <i>l</i>	%	
Anual		5%
Mensal	20%	
Semanal	15%	
Diária	45%	
Intra-Diária (Total)		15%
Intra-Diária	Algumas horas	4%
	De hora em hora	3%
	Até 30 min.	3%
	Até 15 min.	2%
	Abaixo de 1 min.	3%

TABELA 5 - Perfil de Atualização do DW em 2007 (ECKERSON, 2007)

No caso dos processos intra-diário, observa-se uma parcela significativa (8%) de processos com atualização em intervalos inferiores a 30 minutos. Outro aspecto interessante dessa observação é a presença de processos de atualização em intervalos inferiores a um minuto, mesmo que em uma parcela pequena (3%).

# 3.3 Decisões em Tempo Real

#### Segundo Paniam (2007):

"a tomada de decisão em tempo real é um processo que permite às organizações automatizar as suas decisões com base em objetivos e regras pré-definidos. As organizações que optam por esse tipo de abordagem acabam por embutir mecanismos capazes de decidir e executar ações em processos de negócio específicos".

Um aspecto importante é que ao mesmo tempo em que essas organizações automatizam parte do processo decisório elas podem fazer uso das decisões tomadas anteriormente para se aprimorar comparando os resultados obtidos com os esperados para cada um dos modelos de decisão implantados. O processo de decisão em tempo real pode ser descrito como um processo de três etapas, conforme ilustra a figura 9, juntamente com dois exemplos.



FIGURA 9 - A tomada de decisão em tempo real (PANIAM, 2007)

#### Segundo Davenport e Harris (2005):

"Devido às limitações tecnológicas e aos aspectos comportamentais, as primeiras gerações de sistemas especialistas, que surgiram na década de mil novecentos e setenta, não foram bem sucedidas. Nos dias de hoje, entretanto, com uma maior disseminação das tecnologias de apoio a decisão, muitas empresas voltam a empregar essas ferramentas no que se chamou de sistemas de automação da tomada de decisão".

A automação do processo decisório requer o envolvimento de pessoas. O objetivo é permitir que o próprio sistema tome uma decisão ou efetue uma recomendação. Os sistemas de automação da tomada de decisão diferem dos demais sistemas de suporte à decisão em dois aspectos. Em primeiro lugar, em boa parte dos casos não é preciso que uma pessoa reconheça a existência do problema ou então que inicie a análise para que o sistema entre em ação. Ao contrário disso, as decisões são embutidas nos processos de trabalho usuais sem a necessidade de intervenção

humana. O envolvimento das pessoas geralmente é menor, mas não significa que ele seja desnecessário.

Dependendo do tipo de aplicação faz-se necessária a intervenção de uma pessoa para revisar e confirmar a decisão efetuada pelo sistema. Um exemplo comum é a aplicação de tais sistemas na área médica. Em segundo lugar, ao contrário do que ocorre com os sistemas de suporte à decisão, os sistemas de automação da tomada de decisão encontram-se comumente ligados diretamente aos processos de negócio da empresa de modo a transformar rapidamente as decisões em ações. Na maior parte dos casos, esse sistemas têm sido empregados naquelas decisões que ocorrem com alta frequência e com informações que se encontram disponíveis eletronicamente.

O sucesso na adoção de tais sistemas depende da capacidade da organização em codificar o conhecimento dos especialistas em regras e da disponibilidade de dados confiáveis e com qualidade para a automação do processo decisório. Existem muitas aplicações de sucesso que vão desde o controle de sobrecarga nas redes de transmissão de energia elétrica, no diagnóstico e tratamento de doenças, na precificação de produtos e serviços até o cultivo de videiras para a produção de vinho.

#### 3.4 Novas Fontes de Dados

Uma mudança significativa no ambiente de *Data Warehouse* nos últimos anos tem sido uma maior variedade nos assuntos tratados e uma maior diversidade nas fontes de dados. Tradicionalmente, as fontes de dados eram os sistemas transacionais da própria organização. Nos últimos anos, a quantidade de sistemas transacionais cresceu com a difusão da tecnologia da informação em toda a organização. Consequentemente, a quantidade de fontes de dados internas cresceu significativamente. Se, no passado, tradicionalmente, as áreas servidas pelo *Data Warehouse* eram as de vendas e financeira respectivamente, nos dias de hoje, é

possível afirmar que praticamente todas as áreas fazem ou deveriam fazer uso do *Data Warehouse*. Consequentemente, a adoção do *Data Warehouse* em outras áreas como produção, logística, recursos humanos, gestão da força de trabalho, *marketing*, canais de venda, cresceu e continua a crescer (DAVENPORT; HARRIS, 2007).

Com relação às fontes de dados externas pode-se observar o mesmo fenômeno. Nos últimos anos, um número crescente de organizações vêm incorporando dados de fontes externas como mapas, previsão meteorológica, pesquisas de mercado, indicadores de mercado, de redes sociais e até mesmo de sensores em seu ambiente analítico. Estes dados, aliados aos avançados modelos estatísticos (data mining e knowledge discovery), vêm permitindo a essas empresas obterem novos ganhos em seu desempenho.

## 3.5 Desafios Tecnológicos

Dentre os desafios apresentados às arquiteturas de integração de dados na adoção do *Real-Time Business Intelligence*, a redução da latência nos dados continua a ser o mais difícil de superar. A abordagem tradicional para sistemas de *Business Activity Monitoring* considera que essa necessidade é suprida por uma ferramenta capaz de filtrar, limpar, transformar e integrar os dados relevantes provenientes dos ambientes *OLTP* e *OLAP*, assim como, de fluxos de dados (*data streams*). Em muitos casos, dependendo da latência máxima tolerada pelo negócio, é possível utilizar o próprio *Operational Data Store* para esse propósito. Entretanto, em muitos casos, essa abordagem não é viável. Assim sendo, é preciso abandonar o *Operational Data Store* e partir para técnicas que permitam realizar a integração dos dados quase que instantaneamente, como descrito mais a frente.

Uma alternativa bastante estudada nos últimos anos, mas que ainda apresenta algumas limitações, é a utilização das ferramentas de *Enterprise Information Integration* (ou EII). Elas permitem a criação de uma camada de integração virtual e atuam a partir da geração automática de consultas em bases heterogêneas (*query* 

rewriting). Infelizmente, esta abordagem apresenta limitações práticas, principalmente onde os requisitos de latência são rigorosos, quando é necessário manipular eventos ou fluxos de dados, ou então, onde existe a necessidade maior por "limpeza" nos dados. No entanto, ela tem um bom potencial nos casos em que os requisitos de latência não são o fator mais crítico e os dados são estruturados em sua totalidade. A demanda crescente para o uso de fluxos de dados (data streams) pode ser um fator limitador para o EII; entretanto, ainda se está longe de uma solução que possa ser efetiva frente aos diferentes tipos de dados (STONEBRAKER; ÇETINTEMEL, 2005).

O maior desafio segundo Golfarelli et al (GOLFARELLI; RIZZI; CELLA, 2004) parece estar associado à construção de dois componentes: o *Real-Time Integrator* e o *Dynamic Data Store*. Dentre os fatores associados a esses componentes que podem definir a viabilidade de uma tecnologia estão:

- A latência nos dados, definida como a distância entre a ocorrência do evento e a sua percepção pelo usuário, através da alteração de algum indicador;
- A heterogeneidade dos dados, definida como os diferentes tipos de dados: relacional, eventos e fluxos de dados;
- A necessidade de agregação dos dados em tempo real;
- O volume dos dados processados e a necessidade de tratamento de fluxos de dados (ou data streams).

#### 3.6 Níveis de Maturidade

O Real-Time Business Intelligence ou Operational Business Intelligence, como é proposto por Eckerson, engloba um conjunto de práticas, processos e tecnologias.

Sua adoção pode ser classificada em quatro níveis de maturidade ou estágios, conforme a figura 10, segundo Eckerson (ECKERSON, 2007). É importante ressaltar o papel que a latência nos dados possui nesta classificação, uma vez que, quanto menor a latência, potencialmente, existe maior retorno para o negócio a partir da execução efetiva dos processos de negócio.

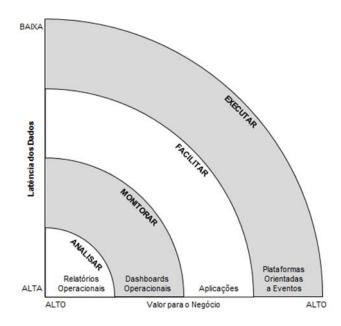


FIGURA 10 - Estágios na adoção do Real-Time Business Intelligence (ECKERSON, 2007)

No primeiro estágio, os usuários fazem uso dos tradicionais relatórios operacionais para analisar seus processos de negócio. Num segundo estágio, os principais usuários conseguem monitorar o negócio através do uso de KPIs, ou *Key Performance Indicators* que são apresentados, seja sob a forma de *dashboards*, seja sob a forma de relatórios operacionais mais elaborados. Num terceiro estágio, a equipe de TI pode facilitar o processo, embutindo os KPI's nas aplicações de missão crítica, facilitando-se assim a análise ao unir tanto operacionais quanto analíticos em uma única aplicação. No último estágio, a empresa é capaz de executar seus processos através do uso de modelos de análise baseados em eventos, de modelos de predição, além de outras técnicas disponíveis para a monitoração e execução de regras de negócio.

#### 3.6.1 Análise de Processos

Nesta etapa, o principal objetivo das organizações é obter uma visão integrada das atividades que são controladas em uma ou mais aplicações, de modo a auxiliar os colaboradores em nível operacional a aprimorarem suas decisões. Neste caso, utilizam-se os tradicionais relatórios operacionais nessa tarefa. Tais relatórios apresentam visões simples das atividades, geralmente, contendo tabelas ou listas, e são utilizados para auxiliar os colaboradores da linha de frente do negócio a tomar decisões de modo a garantir que os processos sejam executados da melhor forma possível. Tais relatórios são tão comuns que antecedem a adoção do *Data Warehouse* nas empresas.

Embora existam muitas alternativas para a implantação dos relatórios operacionais, duas são mais comumente utilizadas: a geração de relatórios a partir de consultas diretamente nos sistemas transacionais ou através de uma base de dados a partir da qual foram transportados os dados dos sistemas transacionais.

A geração dos relatórios operacionais, através de acesso aos sistemas transacionais, pode ser realizada de três formas:

- relatórios, empacotados com sistemas de aplicativos, fazendo acesso direto à base de dados:
- uma ferramenta geradora de relatórios, fazendo acesso diretamente à base de dados:
- 3) ou através de uma ferramenta de EII (*Enterprise Information Integration*) permitindo o acesso a múltiplos sistemas transacionais de uma só vez.

No primeiro caso, os relatórios têm a vantagem de ser fornecidos pelo fabricante da aplicação. Por esse motivo têm a vantagem de exigir menor esforço de desenvolvimento. Embora sejam, comumente, estáticos e pouco flexíveis o que

dificulta a sua adaptação em caso de necessidade. Outra desvantagem de tais relatórios é que, geralmente, eles apenas conseguem servir-se de dados provenientes de uma única aplicação. No segundo caso, as ferramentas de geração de relatórios oferecem uma maior flexibilidade quanto ao acesso a diferentes bases de dados; embora, seja necessário trabalhar com diferentes conjuntos de metadados. Neste caso, o desafio passa a ser a integração dos dados entre sistemas transacionais. No último caso, as ferramentas de EII oferecem a máxima flexibilidade ao permitirem a integração lógica dos metadados. As desvantagens na utilização desse tipo de ferramenta são a complexidade na integração dos metadados e o desempenho na execução de consultas distribuídas.

Um fator de desvantagem em todas as abordagens apresentadas anteriormente é que elas comumente afetam negativamente o desempenho dos sistemas transacionais. Como forma de se evitar isso muitas organizações eliminam a carga de seus sistemas transacionais utilizando um ODS (*Operational Data Store*). Um ponto de desvantagem com relação a essa abordagem é que existem os custos associados à construção e manutenção de tais bases de dados. Se por um lado eles endereçam a questão do impacto do acesso direto nos sistemas transacionais, por outro acrescentam um esforço de construção e manutenção.

# 3.6.2 Monitoração de Processos

Na monitoração de processos, o principal objetivo das organizações é alertar seus colaboradores sobre as condições de exceção e fornecer dados que permitam a eles analisar as respectivas causas. Neste caso, utilizam-se dashboards para o acompanhamento de indicadores de desempenho acerca dos processos de negócio. Existem poucas diferenças entre um dashboard e um relatório operacional. Os dashboards fazem uso das mesmas informações contidas nos relatórios operacionais, com a diferença de que eles comumente enfocam as condições de exceção e apresentam os resultados personalizados para cada usuário. Além disso,

tais mecanismos costumam usar uma gama maior de recursos visuais para facilitar a visualização dos indicadores de negócio.

No caso dos *dashboards*, por utilizarem as mesmas informações que os relatórios operacionais, existem as mesmas questões relacionadas ao acesso aos dados no caso da sua implantação.

### 3.6.3 Facilitação de Processos

Na etapa de facilitação de processos, o principal objetivo é embutir métricas e relatórios operacionais em aplicações existentes e portais. Neste caso, a preocupação primordial é concentrar as informações relativas ao desempenho em poucas ou numa única interface. Através do uso de algumas tecnologias como APIs, XML e *Web-Services*, é possível consolidar informações dos *dashboards* e relatórios nas aplicações de missão crítica ou, em última instância, em portais corporativos. Desta forma, os colaboradores da linha de frente não precisam mais fazer o uso de diferentes ferramentas para realizar o acompanhamento efetivo dos processos de negócio.

# 3.6.4 Execução de Processos

Nesta etapa, o principal objetivo é permitir à organização a captura de eventos de negócio, executar a análise em tempo real de suas causas e ações a serem tomadas, através de um conjunto de regras, e executar as ações necessárias para cada evento, se possível, automaticamente. Neste caso, o que se espera, é utilizar todo o conhecimento da organização a respeito do seu desempenho na execução de seus processos.

A implantação deste tipo de capacidade ainda representa um desafio para a maior parte das empresas, pois nesse caso existe a necessidade de se lidar com dados em baixa latência.

A execução de processos inicia-se a partir da captura dos eventos de negócio, o que geralmente é realizado através da observação de condições ou eventos específicos nos sistemas transacionais, através do uso de *database triggers*, ou a partir de estímulos externos, como, por exemplo, a chegada de uma mensagem.

Uma vez que os eventos são capturados é preciso tratá-los. O tratamento dos eventos pode exigir um processo de análise de dados mais simples podendo ser complexo em outras ocasiões. Nos casos mais simples, os eventos são tratados praticamente de forma isolada, ou seja, cada evento é caracterizado e para esse evento existe um conjunto de regras de negócio a ser aplicado ao mesmo. Um paradigma bastante conhecido com essas características é o BAM (*Business Activity Monitoring*). Neste caso os eventos são tratados de modo a compor indicadores de negócio e em determinadas situações bastante específicas a alteração destes indicadores pode gerar alertas, notificações ou em alguns casos algumas ações mais simples.

Outro paradigma, menos conhecido, é o CEP (Complex Event Processing). Através desse modelo um conjunto de eventos é tratado como um todo de modo a se caracterizar circunstâncias de negócio mais complexas de modo que ações possam ser tomadas. Os sistemas de algorithmic trading, em crescente utilização no mercado de capitais, são exemplos desse tipo de aplicação. Neste caso, esses sistemas recebem dados de mercado através de eventos e tomam decisões e ações sob a forma de ordens de compra e de venda no mercado.

## 3.7 Arquiteturas de Referência

Apresentam-se a seguir duas arquiteturas de referência. No primeiro caso, os autores Golfarelli, Rizzi e Cella propõem uma arquitetura que se pode considerar uma evolução a partir do *Data Warehouse*, acrescentando a esse, alguns componentes para o tratamento de dados de baixa latência. No segundo caso, há o novo paradigma de DW 2.0 proposto por Inmon, Strauss e Neushloss. Neste caso, os autores propõem um novo modelo onde o ciclo de vida e a natureza dos dados têm um papel fundamental.

### 3.7.1 Modelo Proposto por Golfarelli, Rizzi e Cella

Golfarelli, Rizzi e Cella (GOLFARELLI; RIZZI; CELLA, 2004) propõem uma arquitetura geral de acordo com a figura 11.

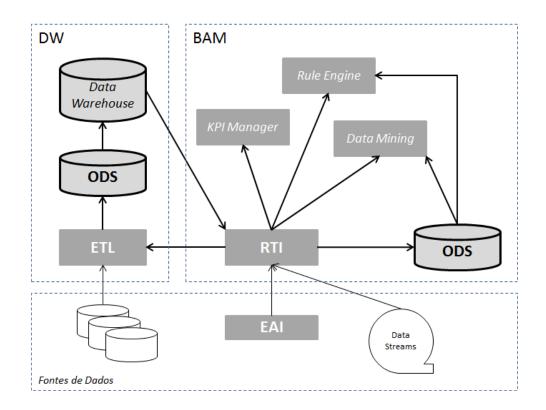


FIGURA 11 - Exemplo de Arquitetura (GOLFARELLI; RIZZI; CELLA, 2004)

Essa arquitetura pode ser vista como uma evolução do *Data Warehouse* (lado esquerdo da figura 11) com uma camada de *ETL* (*Extract-Transform-Load*), responsável pela extração dos dados a partir dos sistemas operacionais, e uma camada *ODS* (*Operational Data Store*), onde são armazenados os dados mais correntes e para uso operacional, e, por final, os dados são carregados no *Data Warehouse*, a partir do ODS, para uso estratégico. Nessa arquitetura, os dados, tanto para fins estratégicos, como operacionais, podem ser acessados a partir de ferramentas *OLAP* ou de geração de relatórios.

O lado direito da figura 11 ilustra os elementos que, segundo os autores, compõem o "fluxo de dados reativo" e que são os responsáveis por permitir a monitoração dos processos críticos em tempo real. O conjunto de tecnologias empregadas na construção dessa competência de monitoração e análise é comumente chamada de *Business Activity Monitoring (BAM)*. Dentre seus elementos é importante mencionar:

- Right-Time Integrator (RTI) que é responsável por integrar dados em tempo real a partir do Data Warehouse, dos sistemas transacionais, dos sistemas de Enterprise Application Integration (EAI) e a partir de fluxos de dados (ou data streams).
- Dynamic Data Store que é responsável pelo armazenamento dos dados por um período curto de tempo, de modo que possam ser recuperados para a realização de data mining, aplicação de regras de negócio e consolidação antes de serem apresentados em um dashboard (por exemplo).
- KPI Manager que é responsável pelo cálculo de todos os indicadores em todos os níveis. Esse elemento é responsável por alimentar os dashboards e cockpits.
- Ferramentas de Data Mining que consigam extrair padrões a partir de fluxos de dados (ou data streams).

Rule Engine capaz de processar os eventos oriundos do RTI ou detectados
pela ferramenta de data mining. Esse mecanismo é o responsável pelo envio
de alertas aos usuários.

Segundo essa proposta, a composição do *Data Warehouse* e *Business Activity Monitoring* pode suprir as necessidades básicas para que uma organização adote o *Business Performance Management* como ferramenta de gestão. O conjunto dessas arquiteturas tem o potencial de oferecer as capacidades de Monitorar, Analisar e Planejar em diferentes níveis, de acordo com as necessidades do negócio.

#### 3.7.2 O DW 2.0

Inmon, Strauss e Neushloss (INMON; STRAUSS; NEUSHLOSS, 2008) propõem uma nova concepção para o *Data Warehouse*. Nesta nova proposição os autores reconhecem a necessidade do tratamento do ciclo de vida dos dados e o perfil de acesso dos mesmos na arquitetura de *Data Warehouse*. Desta forma, um novo modelo é apresentado. Este modelo se baseia em quatro setores. Cada setor é responsável pelo armazenamento dos dados segundo seu perfil de utilização e em função do ciclo de vida dos dados. Os setores propostos são os seguintes:

- Interactive Sector (onde residem os dados correntes, ou seja, aqueles capturados nos últimos 2 segundos);
- Integrated Sector (onde residem os dados capturados nas últimas 24 horas);
- Near Line Sector (onde residem os dados históricos considerando-se um intervalo de retenção de 3-5 anos);
- Archival Sector (onde residem os dados históricos).

Os autores não tratam das questões relacionadas à implantação de cada setor em detalhes; nem mesmo descrevem alternativas de como fazê-lo. Os autores parecem estar mais preocupados em ampliar o conceito de *Data Warehouse* de modo a englobar as novas demandas por dados integrados nas organizações, como aquelas descritas neste trabalho. Em todo caso, o DW 2.0 tem a importância de reconhecer a necessidade da manipulação de dados correntes no ambiente de *Data Warehouse*, o que não se fazia em sua concepção original.

# 4 INTEGRAÇÃO DE DADOS EM TEMPO REAL

A integração de dados em tempo real ainda representa um desafio para a maior parte das organizações que vêm buscando a adoção do *Real-Time Business Intelligence*. Dentre os motivos para isso pode-se mencionar os custos, que não são pequenos, para a transformação de um ambiente analítico que se volte ao tempo real. Além disso, é necessária a utilização de mão-de-obra muito especializada e experiente. Outro aspecto importante é que à medida que a necessidade por dados integrados em tempo real aumenta, também crescem, proporcionalmente, as necessidades por maior confiabilidade, disponibilidade e escalabilidade, ou seja, fatores que também contribuem para um maior custo de entrada.

Ultrapassar as barreiras impostas pela latência nos dados pode exigir investimento e esforços bastante significativos. Uma vez que os dados são capturados ou alterados nos sistemas transacionais, como garantir que, em alguns instantes, os mesmos serão integrados e apresentados aos usuários para a tomada de decisões táticas e operacionais?

Apesar do acesso a um maior poder computacional, a distribuição de dados integrados com alta velocidade ainda representa um problema complexo. A sua resolução depende do uso de uma variedade de tecnologias (SWOYER, 2007). Parte primordial dessa mudança exige que decisões importantes sejam tomadas, envolvendo a adequação do *Data Warehouse* e da infraestrutura de *Business Intelligence* (ECKERSON, 2010). Apresentam-se, a seguir, alguns mecanismos que permitem às organizações transformarem seus ambientes de integração e distribuição de dados de modo a viabilizar a adoção do *Real-Time Business Intelligence*. Apresentam-se também dois casos reais de aplicação encontrados na literatura como forma de ilustrar a aplicação desses recursos em complementação ao *Data Warehouse*.

## 4.1 Captura ou Aquisição dos Dados

Há muitas formas de se efetuar a coleta dos dados a partir de uma fonte de dados para depois populá-los no *Data Warehouse* ou no *Operational Data Store*. Dentre as formas existentes, um fator comum de diferenciação é a latência nos dados. Há as mais comuns que realizam a captura dos dados de forma incremental através da execução de consultas nas fontes de dados utilizando-se um *flag* ou *timestamp* até as formas mais sofisticadas que utilizam algum mecanismo de *Change Data Capture* que é capaz de detectar automaticamente as mudanças. Outra maneira de classificar os mecanismos de atualização está associada à forma de transferência dos dados, que podem trabalhar no modelo *"pull"* onde os dados são trazidos da fonte de dados para o *Data Warehouse* em intervalos fixos de tempo ou no modelo *"push"* onde os dados são empurrados da fonte de dados para o *Data Warehouse*, uma vez que são atualizados na origem. Por final, as diferentes formas de atualização ainda podem ser comparadas ao impacto que oferecem tanto na origem, ou seja, nas fontes de dados, quanto no destino, ou seja, no *Data Warehouse*. O quadro 3 fornece um comparativo entre diferentes formas de captura dos dados:

	Batch	Mini-Batch	Micro-Batch	Em Tempo real
Descrição	Dados são	Dados são	Dados são	Atualizações são
	carregados	carregados de	carregados de	identificadas e
	integralmente ou	modo incremental	modo incremental	automaticamente
	de modo	ao longo de um	mas com	aplicadas no DW
	incremental em	dia (intra-dia)	intervalos de	
	janelas		duração inferior à	
			uma hora	
Latência nos	Diária ou maior	Algumas horas	Alguns minutos	Segundos
Dados		(> 1h)	(< 1h)	(< 3s.)
Captura dos	Consultas	Consultas	Change Data	Change Data
Dados	utilizando filtros	utilizando filtros	Capture	Capture
	( <i>flag</i> ou	( <i>flag</i> ou		
	timestamp)	timestamp)		
Modo de	Pull	Pull	Push e depois Pull	Push
Transferência				
Impacto na	Alto	Médio	Baixo	Baixo
Origem				
Impacto no	Alto	Médio	Médio	Médio
Destino				

QUADRO 3 - Comparativo das diferentes formas de captura dos dados

#### 4.1.1 Mini Batches ou Micro Batches

Os *Mini Batches* e os *Micro Batches* se assemelham ao método tradicional de extração, transformação e carga realizado no ambiente de *Data Warehouse*. Neste caso em particular, a atualização é feita por meio de consultas às fontes de dados e a carga dos dados é realizada de forma incremental. O modo de transferência dos dados é o modelo "pull", onde os dados são puxados das fontes de dados pelo *Data Warehouse*. O conjunto de dados a ser transferido é determinado com o apoio de *timestamps* ou *flags*. Em todos os casos os dados são atualizados muitas vezes ao longo de um dia, o que muda é a frequência. Embora não exista uma classificação

formal para *Mini Batches* e *Micro Batches*, este trabalho adota como critério de diferenciação a latência nos dados sendo que, no caso dos *Mini Batches*, considerase intervalos de atualização de algumas horas, porém, inferior aos processos diários ou semanais, por exemplo, a cada duas ou quatro horas. No caso dos *Micro Batches*, considera-se intervalos de atualização de alguns minutos, porém, inferiores a uma hora, por exemplo a cada quinze ou vinte minutos.

## 4.1.2 Change Data Capture

O Change Data Capture é um mecanismo de integração em tempo real que permite o monitoramento de alterações importantes em uma fonte de dados para então realizar a sua propagação em tempo real até outra fonte de dados destino. Tal tecnologia surgiu como uma evolução (ou nova funcionalidade) das ferramentas de replicação de dados transacionais. Hoje em dia ela está mais difundida, pois os principais gerenciadores de bancos de dados relacionais a implementam.

As primeiras implantações do *Change Data Capture* foram realizadas com o uso de *database triggers*, o que causava muitos problemas de estabilidade nas aplicações monitoradas. Além disso, o esforço para desenvolvimento e manutenção dessas *triggers* exigia bastante esforço. Atualmente, existem implementações sofisticadas que se baseiam na varredura dos *logs* de transação. Por esse motivo elas são pouco intrusivas e não exigem o mesmo esforço de desenvolvimento e manutenção que as *database triggers*.

Uma aplicação natural dessa tecnologia é na caracterização de eventos de negócio a partir de alterações nas fontes de dados. Esta talvez seja a principal e mais promissora aplicação dessa tecnologia.

#### 4.1.3 Trickle Feed

Denomina-se *Trickle Feed* (ECKERSON, 2010) o processo de se atualizar um *Data Warehouse* ou *Operational Data Store* de forma contínua a partir de um *Enterprise Service Bus*. Assim como ocorre com as demais aplicações servidas por um *Enterprise Service Bus*, o *Data Warehouse* (ou o *Operational Data Store*) passa a subscrever um dado conjunto de eventos. Esta abordagem apresenta alguns desafios como o desempenho na atualização, pois não se trata de uma operação realizada em lote. As atualizações no *Data Warehouse* comumente são realizadas em lote para que se obtenha um melhor desempenho. Como forma de contornar o problema, pode-se realizar a alimentação de uma *Staging Area* a partir de *Trickle Feed*, para então realizar-se a atualização do *Data Warehouse* (ou *Operational Data Store*) através *mini-batches* (TAYLOR, 2000).

#### 4.2 Armazenamento e Processamento dos Dados

Apresentam-se a seguir os principais mecanismos para o armazenamento e processamento dos dados.

# 4.2.1 Operational Data Stores

O Operational Data Store continua sendo a principal arquitetura utilizada pelas empresas quando existe a necessidade por dados integrados em tempo real. O Operational Data Store (INMON, 1998) constitui uma base de dados integrada, orientada a assuntos, volátil (tolerando operações de *update*), porém, com dados correntes (diferente do Data Warehouse que armazena dados históricos). Dentre as características importantes a serem consideradas na construção de um Operational

Data Store estão a latência, o volume e a granulosidade dos dados, além do grau de integração frente ao modelo de dados dos sistemas transacionais.

Há quatro classes de *Operational Data Store* (INMON, 2000). Cada uma delas é formada a partir de combinações das características apresentadas anteriormente. Nas classes I e II quase não existe integração dos dados entre fontes distintas, uma vez que, o *Operational Data Store* contém, na prática, uma réplica dos sistemas de origem (transações). Estas classes são as mais comuns por suprirem as necessidades operacionais básicas como a geração de relatórios e consultas *on-line* e, também, devido à sua menor complexidade e custo de implantação. Nas classes III e IV existe uma maior integração dos dados quando comparados aos modelos dos sistemas de origem e sua heterogeneidade. Esta abordagem oferece maior complexidade de implantação e custo. Por este motivo, costumam ser menos comuns. Quando se trata de suprir informações para as aplicações de *Dashboard* e *Scorecard*, como é o caso dos sistemas *Business Performance Management*, faz-se necessária a utilização de um *Operational Data Store* de classe III ou IV.

No caso das classes I e II, a replicação transacional constitui a principal tecnologia utilizada para a movimentação dos dados das aplicações até *o Operational Data Store*. No caso das classes III e IV, faz-se necessária a adoção de ferramentas de *ETL (Extract-Transform-Load)*. No contexto da integração dos dados a partir de fontes heterogêneas, os desafios do *Operational Data Store* são similares aos do *Data Warehouse*. No caso da latência e da granulosidade dos dados, o desafio é muito maior. Com relação à latência, se o mais comum ainda é trabalhar com atualizações diárias no *Data Warehouse*, no caso do *Operational Dara Store* elas podem ocorrer diversas vezes ao dia. No que diz respeito à granulosidade dos dados, pode-se trabalhar com o menor nível possível, ou seja, nos detalhes das transações.

## 4.2.2 Enterprise Information Integration

Enquanto que no *Data Warehouse* e no ODS os dados integrados são materializados (ou persistidos), no caso das ferramentas de *Enterprise Information Integration*, o que ocorre é o que se chama de integração de dados virtual, uma vez que, através da utilização de tais ferramentas, tem-se a impressão que de fato os dados foram materializados quando na verdade não foram. O que tais ferramentas produzem é um novo *schema* virtual, também chamado de *mediated schema*, através do qual os usuários e as aplicações podem executar consultas (BERNSTEIN; HASS, 2008).

As ferramentas de EII possuem um mecanismo de mediação que traduz as consultas realizadas pelos usuários ou aplicações sobre o *mediated schema* em consultas individuais contra as fontes de dados reais. O resultado das consultas individuais é integrado em tempo de execução dando a impressão ao usuário que este está realizando a consulta sobre um *schema* real, ou seja, materializado. Apesar de ainda ser considerada uma tecnologia emergente, muitas organizações já realizaram implantações em grande escala de soluções baseadas nesta tecnologia.

Apesar do EII não ser considerada uma tecnologia que deva substituir o *Data Warehouse* ou o ODS (HALEVY *et al*, 2005), existem outras aplicações importantes para ela. Um dos principais atrativos dessa tecnologia é a possibilidade de combinar, em tempo de execução, os dados de diferentes fontes em ambientes heterogêneos. Apesar do desempenho das consultas muitas vezes não ser o ideal, a abordagem de se utilizar um *mediated schema* pode ser bastante útil em situações de fusões e aquisições. Enquanto não existe tempo hábil para integrar os dados das fontes em um *Data Warehouse*, uma ferramenta de EII pode ser a solução de curto-médio prazo.

### 4.2.3 In-Memory Databases

Pode-se considerar a análise de dados em memória como uma das mais recentes ferramentas de integração de dados em tempo real sendo adotadas pelas grandes empresas. Esta tecnologia faz uso do recente desenvolvimento dos *chips RAM* (*Random Access Memory*), o que possibilita utilizar quantias antes não imaginadas de memória, inclusive em estações *desktop*. Essa nova geração de ferramentas *OLAP* faz uso de estações com processadores de 64-bits que possuem um maior endereçamento de memória e assim permitem integrar, agregar e formatar os dados com maior eficiência.

A figura 12 ilustra a diferença entre o acesso aos dados através de um gerenciador convencional de banco de dados e através de uma ferramenta de análise em memória.

#### CONSULTA EM BASE DE DADOS CONVENCIONAL

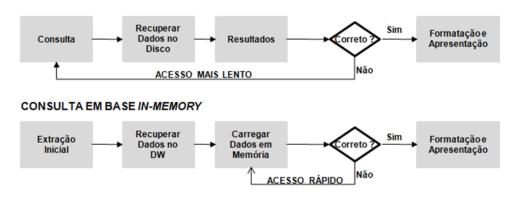


FIGURA 12 - Abordagem convencional versus In-Memory (HOWSON, 2009)

No caso convencional, os dados são recuperados do disco a partir de consultas. Isso requer que os dados estejam previamente integrados em um *Data Warehouse* ou *Operational Data Store*. Os resultados são apresentados ao usuário a partir da execução de cada uma das consultas. A cada nova consulta, é necessário recuperar dados do disco para a memória do servidor. Neste caso, o principal gargalo para a análise dos dados passa a ser o acesso ao disco.

No caso do processamento em memória, um conjunto de visões integradas dos dados é extraído a partir do *Data Warehouse*, do *Operational Data Store* ou até mesmo dos sistemas fonte e carregado em memória. Comumente, utiliza-se alguma técnica de armazenamento eficiente como os *column-stores* ou mecanismos de compactação para otimizar o armazenamento e a recuperação dos dados na memória.

Uma vez que os dados estão em memória, o usuário faz uso da velocidade dos processadores de 64-bits para transformar e integrar os dados sem a necessidade de acesso ao disco. Isto permite uma velocidade superior no acesso a esses dados.

## 4.2.4 Stream Processing

A difusão das chamadas *Stream Processing Applications* cresceu nos últimos anos, principalmente, devido à adoção da tecnologia de sensores nas grandes organizações. Apesar disso, a necessidade por se tratar fluxos de dados nas aplicações surgiu há alguns anos no mercado de capitais, com os sistemas de transmissão eletrônica de transações financeiras (também conhecidos como *Straight Throught Processing ou STP*).

Lidar com fluxos de dados não é tarefa simples, principalmente quando o volume e a velocidade com que novos dados são capturados são altos. O primeiro desafio técnico está associado ao armazenamento desses dados. Na maioria das aplicações, o fluxo tradicional de processamento, também conhecido como "Outbound" processing, considera a captura do dado, armazenamento e posterior processamento. Quando se trata de fluxos de dados, o armazenamento anterior ao processamento pode inviabilizar a aplicação do ponto de vista de desempenho. Isto porque existe um custo de armazenamento que pode ser maior do que o próprio intervalo disponível para o processamento dos dados (sob o ponto de vista da aplicação). Neste sentido, faz-se necessária a adoção de tecnologias que permitam, em primeiro lugar, processar os dados para depois armazená-los (se for realmente

necessário). Esta abordagem é denominada "Inbound" processing. A figura 13 ilustra os dois paradigmas:

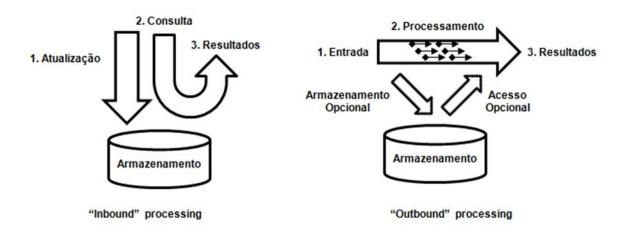


FIGURA 13 - Processamento "Inbound" versus "Outbound" (STONEBRAKER; ÇETINTEMEL, 2005).

A tecnologia a ser utilizada no tratamento de fluxo de dados, do ponto de vista de integração, está diretamente relacionada à máxima latência permitida no processo de análise. Quando a latência é da ordem de minutos, pode-se trabalhar com uma base de dados relacional convencional associada a um mecanismo de captura e atualização dos dados que é semelhante aos *mini-batches*. Neste caso, a base de dados contém fotografias do fluxo de dados de tempos em tempos. Para os casos onde a latência é inferior a poucos minutos, faz-se necessária a adoção de bases de dados especializadas.

A maior parte das aplicações comerciais ainda recai sobre o primeiro caso por trabalhar com latência nos dados da ordem de minutos. Entretanto, começam a surgir novas aplicações, principalmente as relacionadas a sistemas de interação com o cliente, sistemas de controle, segurança ou fraude, que exigem a utilização de tecnologia de bancos de dados voltada a fluxos de dados.

### 4.3 Visualização de Dados

Identificar de maneira rápida as tendências mais importantes para um conjunto de dados pode fornecer uma vantagem, permitindo reduzir a latência de decisão, segundo o modelo da seção 3.2. Quando o decisor reconhece alguma tendência de forma visual pode utilizar todo o ferramental de análise disponível para esclarecer suas suspeitas. As tecnologias visuais disponíveis nos dias de hoje podem condensar muitos dados em uma ou poucas imagens, tornando mais simples o processo de se transformar dados em informação e esta em conhecimento. Por tecnologias visuais considera-se as imagens digitais, sistemas geográficos, interfaces gráficas de usuário, gráficos, representações dimensionais e até vídeos e animações. Este conjunto de recursos, quando combinado, permite uma compreensão mais fácil e ágil de dados e eventos para o usuário final.

A evolução das tecnologias de visualização de dados caminhou em paralelo e teve significativa participação na evolução do *Business Intelligence*, desde o início da década de mil novecentos e noventa. O avanço se deu tanto na computação convencional, onde é integrada às ferramentas e aplicações de suporte a decisão, como nos recursos inteligentes de visualização que incluem a interpretação dos dados. As primeiras ferramentas OLAP forneciam suporte à visualização dos dados apenas em duas dimensões, enquanto que, nos dias de hoje, o suporte a dados em três dimensões é algo bastante comum.

#### 4.3.1 Dashboards

Pode-se dizer que os *dashboards* são a evolução dos *Executive Information Systems* (*EIS*) que surgiram a partir do início da década de mil novecentos e oitenta. Embora o objetivo inicial fosse apresentar informações financeiras através de interfaces simples e de modo que qualquer executivo pudesse compreendê-las, posteriormente, o conceito mostrou-se visionário ou à frente de seu tempo.

Foi necessária uma série de outros desenvolvimentos até se chegar ao conceito de dashboard como é conhecido nos dias de hoje. Pode-se citar o desenvolvimento do Data Warehouse, das ferramentas OLAP e, de certo modo, do amadurecimento do conceito de Business Intelligence na década de mil novecentos e noventa como agentes dessa transformação. Aproximadamente na mesma época, o Balanced Scorecard começa a se difundir como a principal metodologia para a avaliação do desempenho nas organizações. Junto a tudo isso, a necessidade pelo acompanhamento de uma quantidade cada vez maior de indicadores de desempenho. Essa necessidade permitiu que aquela visão inicial dos *EIS* evoluísse até os avançados dashboards disponíveis hoje. Além disto, na virada do século, alguns escândalos como o da Enron resultaram em uma maior pressão para as empresas no sentido de demonstrar para seus acionistas que o seu corpo gerencial realmente tem controle sobre a organização. Este novo fator fez com que as empresas empenhassem esforços na criação de novos controles internos e de meios para permitir que os gestores nos diversos níveis da organização tenham acesso fácil de indicadores de desempenho.

Com base na necessidade de mercado, os fornecedores de *software* de *Business Intelligence* investiram na criação de ferramentas que permitem a construção de *dashboards*. Segundo Stephen Few, no início houve certa confusão entre os diferentes fabricantes de *software* acerca do conceito de *dashboards* (FEW, 2004).

O autor define um *dashboard* como um mostrador visual para as principais informações necessárias para se atingir um ou mais objetivos, consolidadas e arranjadas em uma única tela de modo que a informação seja monitorada de uma só vez (FEW, 2006). Complementando a definição de Few, tem-se:

- Dashboards são mostradores visuais, ou seja, a informação é apresentada de forma visual, através de uma combinação de texto e gráficos, mas com uma maior ênfase em gráficos. A razão pela predominância nos gráficos é que esses elementos permitem uma comunicação mais eficiente do que os textos.
- Dashboards apresentam as principais informações necessárias para se atingir objetivos específicos, ou seja, para se alcançar certos objetivos faz-se

necessário o uso de uma coleção de informações não necessariamente relacionadas que geralmente são obtidas a partir de diversas fontes.

 Dashboards apresentam informações arranjadas em uma única tela de computador, ou seja, as informações necessárias ao se atingir um objetivo devem ser todas traduzidas em uma combinação de texto e gráficos que permitam ao usuário uma comunicação efetiva do que está acontecendo.

Segundo Eckerson (2006), os dashboards devem ser classificados em três tipos:

- Operacionais: São dedicados à monitoração de processos de negócio operacionais, executados pelos colaboradores da linha de frente, que lidam diretamente com o cliente e que são responsáveis pela entrega de produtos ou serviços. Esse tipo de dashboard, geralmente, entrega informação mais detalhada e com pouco grau de sumarização;
- Táticos: São dedicados à processos departamentais ou a projetos de interesse de um segmento ou divisão da organização. Os gestores utilizam este tipo de dashboard para avaliar o desempenho de projetos, do orçamento, de previsões, etc;
- Estratégicos: São dedicados ao acompanhamento e execução dos objetivos estratégicos. Seu objetivo principal é permitir aos gestores do negócio promover o alinhamento das diferentes áreas da empresa em torno dos objetivos maiores a fim de que todas colaborem para que a empresa caminhe no rumo certo.

De acordo com Stephen Few, há muitas variações de *dashboards* conforme ilustra o quadro 4:

Variação	Exemplos
	Estratégico
Papel	Analítico
	Operacional
Tine de dedes	Quantitativo
Tipo de dados	Não Quantitativo
	Vendas
	Financeiro
Domínio dos Dados	Marketing
	Produção
	Recursos Humanos
	Balanced Scorecard
Tipo de Medida	Six-Sigma
	Corporativo
Distribuição dos Dados	Departamental
	Individual
	Mensal
	Semanal
Frequência de Atualização	Diário
	De hora em hora
	Em tempo real
Interatividade	Estático
meratividade	Dinâmico (drill-down, filtros, etc)
	Predominantemente textual
Mecanismos de Apresentação	Predominantemente gráfico
	Balanceado
Intogração com Porteia Cornerativas	Elo de ligação para outros dados
Integração com Portais Corporativos	Sem integração

QUADRO 4 - Tipos de Variação para os Dashboards (FEW, 2006)

O quadro 5 apresenta as diferenças e semelhanças entre os três tipos de dashboards:

	Operacional	Tático	Estratégico
Propósito	Monitorar operações	Medir o progresso	Executar a estratégia
Usuários	Supervisores, especialistas	Gerentes, analistas	Executivos, gerência de alto escalão
Escopo	Operacional	Departamental	Corporativo
Informação	Detalhada	Detalhada e Sumarizada	Detalhada e Sumarizada
Atualização	Intra-diária	Diária e semanal	Mensal, Trimestral
Ênfase	Monitoração	Análise	Gerenciamento

QUADRO 5 - Os três tipos de dashboards (ECKERSON, 2006)

#### 4.3.2 Sistemas de Informação Geográfica

Um sistema de informação geográfica, ou GIS, acrônimo para a mesma expressão em inglês, pode ser definido como um sistema de informação utilizado na captura, armazenamento, modelagem, recuperação, integração, manipulação, análise e exibição de dados geográficos, representados através de mapas digitais. Nestes sistemas, cada registro ou objeto tem uma localização geográfica identificada e integra mapas aos bancos de dados orientados espacialmente a outros bancos de dados de modo que seus usuários possam utilizar tais informações para planejamento, análise e resolução de problemas. Com o passar do tempo, tais ferramentas tornaram-se cada vez mais avançadas e acessíveis e, por esses motivos, a sua utilização no processo de decisão aconteceu de forma natural de modo que é possível, nos dias de hoje, fazer uso de um mapa, como parte de um dashboard gerencial ou operacional. São inúmeras as aplicações e os casos práticos da utilização de mapas associados a dashboards. Da mesma forma, é crescente a utilização de mapas em processos analíticos, tais como a previsão de comportamento de clientes e a prevenção e combate ao crime e ao terrorismo.

Existe grande potencial de aplicações desses sistemas associados à tecnologia de GPS para a localização e rastreamento de recursos como frotas de caminhões, segurança pessoal e patrimonial.

#### 4.4 Casos Reais de Aplicação

A adoção de arquiteturas de integração de tempo real nas grandes empresas ainda é incipiente. Dentre os motivos para isso pode-se mencionar os custos, que não são pequenos, para a transformação de um ambiente analítico que se volte ao tempo real. Além disso, é necessária a utilização de mão-de-obra muito especializada e experiente. Outro aspecto importante é que à medida que a necessidade por dados integrados em tempo real aumenta, também crescem, proporcionalmente, as necessidades por maior confiabilidade, disponibilidade e escalabilidade, ou seja, fatores que também contribuem para um maior custo de entrada.

A seguir, apresentam-se dois casos reais de aplicação com o objetivo de mostrar a aplicação prática de alguns dos conceitos aqui expostos.

#### 4.4.1 Quicken Loans

A *Quicken Loans* é uma das maiores empresas de empréstimos e hipotecas, nos Estados Unidos, especializada na venda on-line. Seus analistas estão preparados para atender os clientes a partir de qualquer canal eletrônico: telefone, *email* ou *Web*. Apoiando esta operação, está um *Web Call Center* com mais de quinhentas posições de atendimento, onde cada analista tem acesso, em tempo real, às informações dos clientes, do mercado financeiro, do seu desempenho e de seu grupo. Além disso, espalhados pelo salão desse *Web Call Center* existem grandes televisores que mostram, a cada dois minutos, o ranking das quinze principais empresas do ramo, segundo doze indicadores de desempenho.

Além do *Web Call Center*, a empresa ainda conta com um centro de controle por meio do qual os gerentes podem monitorar o desempenho da operação em tempo real (ECKERSON, 2006).

Para suportar a sua estratégia de *Business Performance Management*, a *Quicken Loans* investiu, por dois anos, na criação de uma nova arquitetura de *Business Intelligence*. Ela é responsável por prover uma variedade de dados em tempo real para os diversos níveis da organização, ou seja, desde os atendentes do *Web Call Center* até os executivos.

Antes dessa nova arquitetura, o fluxo de informações envolvia trabalho manual e o uso dos sistemas transacionais da empresa, o que ocasionava também muitos problemas como o baixo desempenho na obtenção das informações sob a forma de relatórios. Diante de mercados dinâmicos, como é o caso da *Quicken Loans*, isto é inviável. Ao contrário do que acontecia anteriormente, agora a empresa consegue alimentar uma força de trabalho composta por, aproximadamente, duas mil e quinhentas pessoas com dados integrados em tempo real (LOFSTROM, 2004).

A arquitetura adotada pela *Quicken Loans* é composta de quatro camadas: 1) Sistemas Fonte; 2) Integração de Aplicações; 3) Integração e Distribuição de Dados e 4) *Dashboards* Táticos e Operacionais, conforme a figura 14.

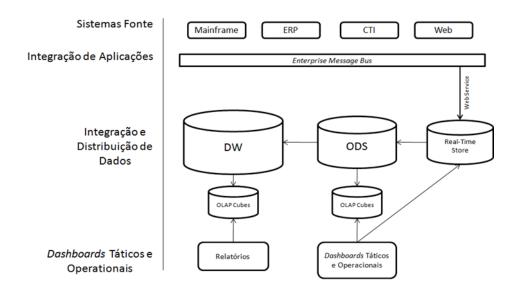


FIGURA 14 - Arquitetura da Quicken Loans (ECKERSON, 2006)

O fator de diferenciação desta arquitetura frente às usuais, nas quais se trabalha, exclusivamente, com *Data Warehouses e Data Marts*, é a adoção de tecnologias ou técnicas especiais na camada de integração e distribuição de dados. A partir da figura 14, nota-se o uso de *Tricke-Feed* para a retirada dos eventos de negócio a partir de um *Enterprise Service Bus*, a adoção de um *Dynamic Data Store* como proposto na seção 3.7.1, além do uso de *Operational Data Store* para o acesso aos dados operacionais.

A aquisição dos dados começa nos sistemas de origem, onde os mecanismos de Change Data Capture notificam os eventos de negócio ao Service Bus que, por sua vez, alimenta o Dynamic Data Store usando a técnica de Tricke-Feed. O Dynamic Data Store que, neste caso, consistiu de uma base relacional convencional ajustada para operar com praticamente a totalidade dos dados em memória, armazena os dados relativos a eventos gerados nos últimos dois dias. Esta base de dados serve como fonte de consultas, além de permitir a realização de operações de limpeza, transformação e agregação de dados. Deste momento em diante, os dados fluem de modo a alimentar o Operational Data Store e o Data Warehouse sucessivamente. Algumas centenas de cubos permitem que os dashboards acessem os dados integrados - agregados ou detalhados - a partir do Dynamic Data Store, do Operational Data Store ou do Data Warehouse.

O Operational Data Store armazena um volume de dados correspondente aos últimos dois meses e o Data Warehouse um volume correspondente aos últimos sete anos.

#### 4.4.2 Yahoo!

O Yahoo! oferece conteúdo e anúncios a seus consumidores através de dezenas de milhares de servidores espalhados pelo globo por meio dos quais são entregues serviços como o Yahoo! Mail, Finance, Messenger, Search e Shopping.

Essa infraestrutura de servidores oferece, além das inúmeras páginas com conteúdo, uma série de eventos contendo informações como o conteúdo, o *layout* 

das páginas, além de dados de interação com o consumidor; tais como o conteúdo visualizado, palavras-chave e *links*.

A coleta, armazenamento e análise desses dados constitui uma das atividades chave da empresa. Para isso, ela implantou uma arquitetura baseada no conceito de linhas de montagem para realizar a aquisição e integração de seus dados. Existe uma linha de montagem para cada grande área de negócio como busca, anúncios e audiência. Cada uma das linhas de montagem conta com um *Operational Data Store* para armazenar dados recentes através da execução de *mini-batches* cujo intervalo varia de cinco minutos a algumas horas. As aplicações que demandam dados com uma latência inferior a dez segundos buscam os mesmos diretamente no sistema de coleta de dados.

Esse sistema de coletas é composto por uma base de dados operando similarmente ao *Dynamic Data Store*, proposto anteriormente. Esta base de dados é alimentada através da execução de *mini-batches* em intervalos muito curtos (segundos) de modo a capturarem os dados coletados nos *logs* de navegação dos *Web Servers*. Ao final do ciclo de vida, os dados detalhados são armazenados em um *Data Warehouse*. Ao longo de um dia, esse sistema é responsável por capturar trinta bilhões de eventos de negócio, o que acarreta um acréscimo de trinta *Terabytes* no *Data Warehouse*. A figura 15 apresenta a arquitetura utilizada no *Yahoo!*.

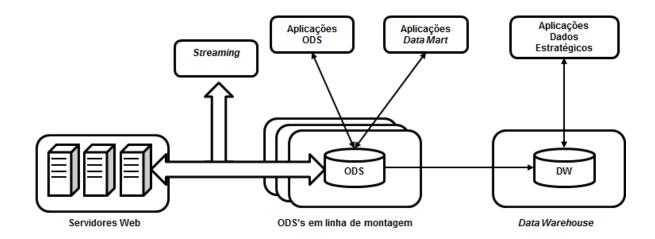


FIGURA 15 - Arquitetura de Integração de Dados em Tempo Real do Yahoo! (AHUJA et al, 2009)

O acesso aos dados é realizado através de três categorias de aplicações: 1) Streaming; 2) Operational Data Store e 3) Data Marts.

As *Streaming Applications* fazem uso de baixíssima latência e, por esse motivo, acessam os dados diretamente do mecanismo de coleta. Estas aplicações, por sua vez, também fazem uso de bancos de dados configurados para tratar os dados em memória. Um exemplo dessas aplicações é aquela que controla o consumo de anúncios nas páginas a partir do acesso dos usuários.

As aplicações que acessam ao *Operational Data Store* são aquelas que consomem dados de baixa latência como aqueles atualizados entre os últimos cinco minutos e o último dia. As aplicações que demandam dados detalhados são construídas com acesso direto ao *Operacional Data Store*. Aquelas que exigem um maior grau de agregação nos dados fazem uso de *Data Marts* construídos acima do *Operational Data Store* ou do *Data Warehouse*. O *Data Warehouse* armazena os dados históricos (mais velhos do que um dia).

No caso do *Yahoo!*, a arquitetura adotada apresentou algumas vantagens. Em primeiro lugar, a empresa tem o desafio de lidar com volumes de dados gigantescos, além de fazer uso de dados integrados com latência variando de alguns segundos a alguns dias. A adoção de diferentes tecnologias permitiu, assim como no caso da *Quicken Loans*, administrar a complexidade imposta pela baixa latência nos dados. De outro lado, criação de linhas de montagens por área de negócio permite uma melhor segmentação dos dados de modo a reduzir o volume e facilitar a administração. Outro benefício é a independência, em termos de infraestrutura, para cada uma das áreas de negócio.

#### 4.5 Considerações

Os casos reais de aplicação exemplificam a utilização dos novos mecanismos de integração de dados, aplicados em diferentes contextos, com o objetivo de promover a inteligência de negócios tempo real. A adaptação do ambiente de Data Warehouse para a redução das latências requer a adoção de mecanismos complementares, como os apresentados neste capítulo. Os casos também servem como ilustração das tendências apresentadas no capítulo três como o enfoque operacional, a redução de latências, as decisões em tempo real, as novas fontes de dados (caso Yahoo!). O caso Quicken Loans ilustra a utilização de mecanismos como o Trickle Feed, o In-Memory Database e o ODS para atender os requisitos de latência de dados e de análise. O caso Yahoo! ilustra a utilização de Stream Processing em conjunto com o ODS e o Data Warehouse para o processamento de dados com baixíssima latência. No primeiro caso, o principal desafio é prover dados integrados a um grande volume de usuários enquanto que no segundo caso o principal desafio é processar um volume enorme de dados. Os dois casos servem para ilustrar a importância do ciclo de vida dos dados, como sugere o DW 2.0 (INMON; STRAUSS; NEUSHLOSS, 2008).

## 5 A ADOÇÃO DA INTELIGÊNCIA EMPRESARIAL EM TEMPO REAL

Quando uma empresa opta por investir na adoção do *Real-Time Business Intelligence* existem desafios técnicos e desafios de negócios a serem endereçados. Como o investimento necessário pode ser significativo, algumas empresas optam por uma abordagem focada em resultados, ou seja, analisam as oportunidades de ganho caso a caso e realizam o investimento através de uma série de projetos. Outras empresas, por sua vez, partem para estratégias mais agressivas e fazem grandes investimentos em sua infraestrutura de tecnologia da informação num primeiro momento para depois colher os benefícios, oportunidade a oportunidade, posteriormente.

O que alguns autores sugerem, como é o caso de McGee, é que as empresas interessadas em se transformar em empresas que realizam negócios em tempo real devem investir em processos internos complementares ao planejamento estratégico para que se definam eventos, métricas e ações que possam ser monitorados, analisados e executados de modo a produzir ganhos para a empresa.

#### 5.1 Reduzindo as Latências

O principal objetivo a ser atingido na construção de uma infraestrutura para a adoção do *Real-Time Business Intelligence* é o de reduzir as latências (dados, análise e decisão) para os processos definidos como alvo dessa iniciativa.

Reduzir a latência nos dados implica, geralmente, em uma revisão de todos os processos, da captura ou aquisição dos dados, passando pelo armazenamento e processamento até a visualização dos mesmos. Os resultados desejados podem ser alcançados por meio da otimização da participação de todos esses elementos ou de alguns isoladamente.



FIGURA 16 - Os elementos que compõe a latência nos dados

A redução na latência da análise pode ser realizada de três formas:

- Oferecendo ao tomador das decisões as informações relevantes, em tempo adequado, para que este possa se concentrar exclusivamente na análise dos dados;
- Criando mecanismos que auxiliem o tomador de decisões, seja acelerando a análise ou sugerindo resultados para a decisão;
- Automatizando o processo decisório.

A redução na latência na decisão, geralmente, é alcançada através da automatização do processo decisório, comumente efetuado através da criação de sistemas de automatização de decisões.

## 5.2 O Modelo de Identificação e Justificação de McGee

Segundo McGee (2004b), algo em torno de 5% das informações utilizadas na condução dos negócios de uma empresa precisa ser monitorada em tempo real. Como a monitoração de dados em tempo real, comumente, representa um investimento em tecnologias avançadas que, por sua vez, acarreta custos, o autor sugere a adoção de um modelo baseado na identificação, ou seja, na definição das informações que, quando monitoradas em tempo real, têm o potencial de trazer benefícios à organização e na justificação, ou seja, se os benefícios possíveis de serem alcançados superam os custos.

A figura 17 apresenta o modelo de identificação de McGee que permite identificar, priorizar e avaliar os eventos cuja monitoração pode trazer benefícios ao negócio.

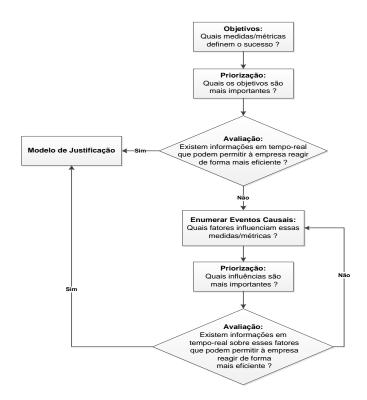


FIGURA 17 - O modelo de identificação de McGee (McGEE, 2004)

O modelo de justificação, ilustrado na figura 18, baseia-se na idéia de funil, onde o conjunto de informações em tempo real definido no processo de identificação é filtrado utilizando-se quatro perguntas importantes.



FIGURA 18 - O modelo de justificação de McGee (McGEE, 2004)

O quadro 6 ilustra um exemplo de modelo de identificação e justificação para uma empresa de transporte aéreo:

Modelo de Identificação		
- Wodelo de Idelitilicação		
Meta	Lucro Operacional	
Eventos Causais	Receita a partir de viagens de negócios e a situação	
	econômica (mais ou menos favorável às viagens de	
	passeio)	
Métricas	Viagens a negócios por segmento de mercado	
Respostas a variações nas métricas	Renegociação de contratos, em particular, os relacionados	
	à compra de combustíveis, postergação de novos	
	investimentos, etc.	
Modelo de Justificação		
Alinhamento Estratégico	Sim. O controle rigoroso do lucro operacional é um fator de	
	sucesso nessa indústria.	
Priorização	Sim.	
Materialidade	Sim. A pronta resposta pode evitar perdas significativas.	
Impacto	Sim. Pode evitar prejuízos e a necessidade de se realizar	
	cortes de pessoal.	

QUADRO 6 - Exemplo de modelo de identificação e justificação para uma empresa do setor de transporte aéreo.

## 5.3 A Visão do Projeto

Na construção da visão para o projeto devem ser definidos os processos e subprocessos de negócio com potencial para o uso do *Real-Time Business Intelligence*. Pode-se aplicar o modelo de McGee na identificação e justificação daquelas informações cujas latências deverão ser reduzidas. Outro aspecto importante, é estabelecer, para cada informação selecionada, o alcance que será dado em função do modelo de maturidade apresentado na seção 3.6.



FIGURA 19 - Estabelecendo uma visão para o Real-Time Business Intelligence

Processos que necessitam de um aprimoramento na análise são aqueles em que há pouca ou nenhuma evidência sobre o seu comportamento. Nos processos onde é necessário um aprimoramento de sua monitoração, a organização já possui conhecimento suficiente para tomar determinadas ações em cada um dos eventos de negócio mapeados. Por último, estão os processos que já existe conhecimento sobre as ações a serem tomadas para cada um dos possíveis eventos, a organização já possui alguma capacidade de detecção desses eventos, mas precisa aprimorar sua capacidade de ação. Para cada processo, pode-se definir um objetivo em termos de maturidade a se atingir: monitoração, facilitação ou execução.

A análise inicial dos processos permite avaliar o grau de maturidade requerido para a solução a ser construída ao mesmo tempo em que serve de apoio na definição da arquitetura de sistemas que irá suportá-los.

A quadro 8 ilustra os objetivos a serem atingidos a partir do grau de maturidade no gerenciamento dos processos, segundo o modelo proposto por Paniam.

Expectativa ou Maturidade	Objetivos de Negócio
Análise	Compreender os eventos de negócios e respectivos "insights".
Monitoração & Facilitação	Tomar ações, através de pessoas, para eventos de negócios e "insights" conhecidos.
Execução	Tomar ações, através de mecanismos, para os eventos de negócio e "insights" conhecidos para os quais a Latência de Ação ou Decisão seja pequena.

QUADRO 7 - Objetivos de Negócio a partir da maturidade dos processos (PANIAM, 2007)

# 5.4 Os Investimentos Necessários na Infraestrutura de Tecnologia da Informação

Segundo Eckerson, em pesquisa realizada pelo *TDWI*, em fevereiro de 2007 (ECKERSON, 2007), da qual participaram empresas membro do TDWI, fornecedores de soluções e especialistas do mercado de *Data Warehouse*, a principal preocupação na adoção do *Real-Time Business Intelligence* está concentrada em: 1) definição de uma infraestrutura robusta, 2) gestão das expectativas, 3) desempenho na execução de consultas para obtenção das informações e 4) na obtenção de recursos (orçamento).

O principal investimento, quando se opta pela adoção do *Real-Time Business Intelligence*, está associado à redução da latência nos dados (segundo modelo proposto por Hackathorn). Os custos de implantação tendem a aumentar à medida que os requisitos por baixa latência nos dados se tornam mais rigorosos. Em muitos casos, os benefícios para o negócio tendem a ser maiores quando os eventos relevantes são capturados e tratados em curtíssimos espaços de tempo. Por esse motivo, é importante que se tenha uma visão bastante clara dos benefícios esperados antes do investimento na infraestrutura para se reduzir a latência nos dados. O uso de modelos como o proposto por McGee podem auxiliar no direcionamento de investimentos. A figura 20 ilustra o acréscimo nos custos e o potencial valor para o negócio à medida que é preciso suportar uma menor latência nos dados:



FIGURA 20 - Latência dos dados versus os custos e o valor para o negócio (ECKERSON, 2007)

#### 5.5 Os Desafios Técnicos

A tarefa de se reduzir a latência nos dados geralmente apresenta alguns desafios técnicos a maior parte das empresas. Uma das primeiras decisões que a equipe técnica precisa tomar é a de avaliar o quanto o seu ambiente de *Data Warehouse* e de *Business Intelligence* estão preparados para aos novos requisitos de latência.

## 5.5.1 O que fazer com o Data Warehouse?

A tendência natural é considerar como melhor opção a reutilização do *Data Warehouse* existente, uma vez que, ele já contém boa parte dos dados integrados da organização. Um dos principais argumentos para isso é que não se deve separar os dados operacionais ou táticos daqueles estratégicos para que não se amplifiquem ou piorem questões difíceis de lidar como a qualidade ou integridade dos dados. Por outro lado, em alguns casos, o *Data Warehouse* pode se tornar um gargalo, uma vez que, mais dados sejam carregados nele em intervalos cada vez menores. A concorrência por processos de atualização e consulta pode comprometer o

desempenho do *Data Warehouse*, o que pode colocar em risco a disponibilização de dados operacionais com baixa latência na integração dos dados.

#### 5.5.2 As Limitações do Ambiente de Data Warehouse Tradicional

A primeira pergunta a se responder, quando se analisa uma arquitetura de integração e distribuição dos dados, é até que ponto pode-se contar com o *Data Warehouse* como a solução para a entrega de dados integrados com baixa latência. Existem muitos métodos que podem ser utilizados na atualização do ambiente de *Data Warehouse* para que se reduza a latência nos dados. Tais métodos podem ser classificados em três categorias:

- Aquisição de Dados: São as ferramentas responsáveis por capturar, transformar e mover os dados dos sistemas de origem até o Data Warehouse.
   Dentre as opções existentes pode-se citar: ferramentas de ETL (Extract-Transform-Load) convencionais, ferramentas de replicação de dados transacionais, change data capture, ferramentas de EAI (Enterprise Application Integration) e ferramentas de EII (Enterprise Information Integration);
- Armazenamento dos Dados: São as diversas camadas ou componentes de armazenamento tais como o banco de dados relacional convencional, staging areas, bancos de dados multidimensionais, bancos de dados de baixa latência ou bancos de dados em memória;
- Entrega ou Apresentação dos Dados: São as ferramentas de apresentação dos dados tais como portais, dashboards, relatórios, alertas, etc.

O maior problema a se enfrentar quando o *Data Warehouse* é utilizado como principal componente na arquitetura de integração e distribuição de dados em tempo real é conciliar os processos de atualização dos dados e as consultas sobre os

mesmos. Neste caso, muitas das técnicas de otimização de desempenho utilizadas no ambiente de *Data Warehouse* tradicional, como eliminar os índices nas tabelas durante os longos processos de carga para depois recriá-los, não podem ser utilizadas devido ao requisito de baixa latência. Consequentemente, paga-se o preço da execução dos processos simultaneamente, uma vez que, não é possível separar carga (ou atualização) das consultas. Para piorar a situação, geralmente exigem-se baixíssimos tempos tanto para a execução dos processos de atualização (quando se utilizam *mini-batches*, por exemplo) quanto para a execução de consultas operacionais.

Dentre as alternativas para se lidar com *workloads* mistos (consultas e atualizações) tem-se:

- Supressão de consultas mais pesadas como forma de se reduzir a sobrecarga no sistema: Essa alternativa pode ser adotada através do agendamento das consultas mais pesadas para períodos de menor carga no sistema (de madrugada, por exemplo) ou através da utilização de recursos de gerenciamento nos bancos de dados relacionais como resource governors. O problema com esta abordagem é que ela fere um dos princípios do Data Warehouse que é o de oferecer acesso à informação sem a sobrecarga dos sistemas transacionais;
- Separar consultas a partir de seu perfil (analíticas ou táticas): Esta abordagem considera uma redução da carga no Data Warehouse através da execução das consultas táticas em um ODS (Operational Data Store) ou Data Mart. Neste caso o ODS ou Data Mart contém os dados mais recentes (os últimos 30 dias, por exemplo) enquanto que o Data Warehouse permanece com os dados históricos. O problema nesta abordagem acontece quando é preciso trabalhar tanto com dados recentes quanto com dados mais antigos;
- Reduzir a contenção nos processos de carga: Quando consultas e processos de atualização são executados sobre as mesmas tabelas existe o risco de um processo bloquear o outro. Há duas formas de se evitar o problema:

- Permitir leituras e escritas simultaneamente: Neste caso devem ser evitadas as operações de atualização em tabelas, pois corre-se o risco de bloqueá-las. Posto isto, as consultas podem ser executadas sem contenções. Como consequência, porém, existe o problema do sincronismo dos dados, já que uma mesma consulta pode obter resultados bem diferentes se executada repetidamente em pequenos intervalos de tempo. Como forma de minimizar estes efeitos, pode-se utilizar filtros nas consultas que contenham algum tipo de timestamp em suas cláusulas, evitando-se assim, grandes variações nos resultados;
- o Insert and Flip de partição: Essa técnica consiste no particionamento das tabelas fato e na utilização de tabelas intermediárias com as mesmas características (mesmas colunas, atributos, etc). Essas tabelas intermediárias sofrem a atualização enquanto que as consultas são realizadas na tabela fato. De tempos em tempos, em intervalos preferencialmente curtos de tempo, os dados da tabela intermediária são transferidos para a tabela fato. Essa técnica é tão utilizada que os principais bancos de dados relacionais de mercado oferecem recursos específicos para a sua implantação.

## 5.6 Selecionar as Tecnologias Adequadas

Assim como apresentado no Capítulo 4, existe um conjunto grande de opções tecnológicas a serem aplicados na construção de uma infraestrutura de integração de dados em tempo real. Assim como no caso do ambiente de *Data Warehouse*, essas tecnologias podem ser classificadas ou organizadas em três categorias:

 Aquisição dos Dados: É preciso capturar, transformar e movimentar os dados a partir de sistemas de origem ou fontes de dados até o ambiente analítico em tempo real. Para essa tarefa, existe uma variedade de ferramentas tais como as tradicionais ferramentas de ETL, de replicação de dados, de *change data capture*, de mensageria (ou *message bus*), ferramentas para o tratamento de eventos e regras de negócios, ferramentas de EII, dentre outras. Na prática, o que ocorre é a combinação de tais ferramentas;

- Armazenamento dos Dados: É preciso armazenar os dados capturados de alguma forma, preferencialmente através de algum gerenciador de banco de dados em alguma camada de integração de dados como as staging area, o Data Warehouse, Data Marts, um ODS ou algum banco de dados de baixa latência ou ferramenta de tratamento de eventos;
- Entrega ou Apresentação dos Dados: A apresentação dos dados capturados e armazenados requer a utilização de aplicações existentes, portais, dashboards, ferramentas de relatórios ou uma combinação destas opções.

Algumas das tecnologias já citadas têm sido mais empregadas do que outras. A figura 21 ilustra as principais ferramentas utilizadas, em grau de importância, a partir de uma pesquisa realizada com 255 empresas (ECKERSON, 2007):

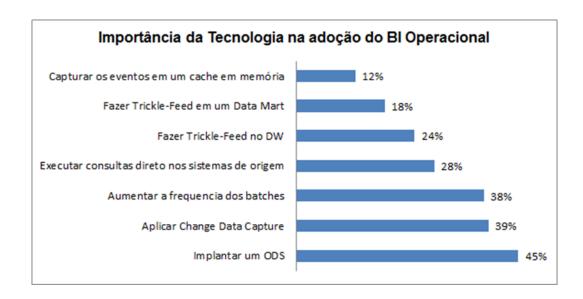


FIGURA 21 - A importância de algumas tecnologias na adoção do BI Operacional, ou *Real-Time Business Intelligence* (ECKERSON, 2007)

## 5.6.1 A Adoção Gradual de Novas Tecnologias

Uma boa estratégia para as organizações que ainda não têm claramente os principais objetivos e benefícios esperados para a tomada de decisão em tempo real pode ser a adoção gradual dessas tecnologias. Assim como ilustra a figura 21, as organizações podem reduzir a latência na integração de seus dados aos poucos, ou seja, à medida que estes sejam utilizados. Por exemplo, uma organização pode iniciar o processo com a construção de um *ODS* extraindo as informações a partir dos sistemas de origem a cada duas ou três horas e, posteriormente, começar a reduzir esse intervalo até utilizar um mecanismo de *Change Data Capture* em alguma de suas fontes. À medida que a demanda por eventos cresce, a empresa pode investir em uma plataforma de gerenciamento de eventos e mensagens e, num momento seguinte, adotar a técnica de *Trickle-Feed* para alimentar seu *ODS* e o seu *Data Warehouse*. São muitas as opções. A principal vantagem desta abordagem é o

investimento em etapas. Uma possível desvantagem é que a organização pode deixar de realizar parte dos benefícios de negócio, ou então postergá-los, devido a essa abordagem mais conservadora.

Apresentam-se a seguir algumas abordagens, baseadas no conhecimento das tecnologias ou em casos reais de aplicação, para a construção de uma infraestrutura para o *Real-Time Business Intelligence*.

## 5.7 A Captura e Aquisição dos Dados

A escolha das tecnologias de aquisição de dados pode variar em função da natureza dos dados e da latência nos mesmos. Comumente, o que ocorre é a utilização de um conjunto de tecnologias. O quadro 9 apresenta as principais tecnologias disponíveis e suas características:

Tecnologia	Principais Características	Latência dos Dados
Mini-Batches	Pode ser utilizada na captura	A frequência dos <i>batches</i> ou intervalo
	dos dados a partir de arquivos	pode ser reduzida de dezenas de
	ou outras bases de dados. O	minutos para alguns minutos desde
	processo de captura dos dados	que o volume de atualização não seja
	ocorre em lotes de registros.	grande. O mais comum é trabalhar
		com atualizações na ordem de
		dezenas de minutos.
Change Data Capture	Pode ser utilizada na captura	A latência geralmente é inferior a um
	de dados com baixíssima	minuto (da ordem de segundos)
	latência a partir de bases de	
	dados. Geralmente, o processo	
	de captura ocorre evento a	
	evento.	
Trickle-Feed	Pode ser utilizada na captura	A latência geralmente é inferior a um
	de dados com baixíssima	minuto (da ordem de segundos)
	latência a partir de um	
	Enterprise Service Bus.	
Stream Application ou	Pode ser utilizada na captura	A latência geralmente é inferior a um
Stream Processing	de fluxos de dados (streams)	minuto (da ordem de segundos) e o
Engine	diretamente de uma aplicação	volume de eventos a serem tratados
	que trate streams de dados ou	de uma só vez é bastante grande
	utilizando-se para isso algum	(superior a milhares por segundo).
i	Stream Processing Engine	

QUADRO 8 - Opções tecnológicas para a aquisição de dados

#### 5.8 O Armazenamento e Processamento dos Dados

A escolha das tecnologias de armazenamento e processamento dos dados geralmente depende de alguns fatores como:

- 1) O perfil de utilização dos dados no processo decisório;
- 2) O ciclo de vida e o armazenamento dos dados.

O perfil de uso dos dados no processo decisório não é uniforme. Ele pode variar em função do estágio, em termos de maturidade: Análise, Monitoração ou Execução.

No processo de análise, o perfil de utilização dos dados se assemelha ao perfil tradicional do *Data Warehouse* onde o principal objetivo é compreender o negócio e buscar "insights" que possam trazer alguma vantagem ao negócio. Apesar dessas semelhanças, nos dias de hoje existem novas necessidades a serem atendidas. Devido a um ambiente de negócios mais competitivo, complexo e dinâmico, as organizações precisam adquirir uma capacidade mais robusta de análise. Por conta disto, novas demandas surgem, tais como:

- Maior variedade nos assuntos e fontes de dados;
- Armazenamento e processamento de um volume dados maior;
- Suporte a um número maior de usuários;
- Uso de avançadas técnicas de data mining e knowledge discovery na obtenção de vantagem competitiva.

#### 5.8.1 Major Variedade nos Assuntos e Fontes de Dados

Uma mudança significativa no ambiente de *Data Warehouse*, nos últimos anos, tem sido uma maior variedade nos assuntos tratados e uma maior diversidade nas fontes de dados. Tradicionalmente, as fontes de dados para o *Data Warehouse* eram os

sistemas transacionais. Nos últimos anos, a quantidade de sistemas transacionais cresceu com a difusão da tecnologia da informação em toda a organização. Consequentemente, a quantidade de fontes de dados internas cresceu significativamente. Se, no passado, as áreas servidas pelo *Data Warehouse* eram as áreas de vendas e financeira respectivamente, nos dias de hoje é possível afirmar que todas as áreas fazem ou deveriam fazer uso do *Data Warehouse*. Consequentemente, a adoção do *Data Warehouse* em outras áreas como produção, logística, recursos humanos, gestão da força de trabalho, marketing, canais de venda cresceu e continua a crescer (DAVENPORT; HARRIS, 2007).

Com relação às fontes de dados externas, pode-se observar o mesmo fenômeno. Nos últimos anos, um número crescente de organizações vêm incorporando dados de fontes externas como mapas, previsão meteorológica, pesquisas de mercado, indicadores de mercado, de redes sociais e até mesmo de sensores em seu ambiente analítico. Tais dados, aliados aos avançados modelos estatísticos (data mining e knowledge discovery), vêm permitindo a essas empresas obterem novos ganhos em seu desempenho.

## 5.8.2 Armazenamento e Processamento de um Volume de Dados Maior

O aumento significativo no volume de dados a ser armazenado e processado se deve a uma série de fatores, tais como:

- A maior diversidade de assuntos tratados;
- Questões regulatórias que exigem mais controles e retenção de dados históricos por períodos maiores;
- A crescente demanda por mais granulosidade nos dados impulsionada por técnicas avançadas de análise de dados;
- A aplicação de técnicas avançadas de análise dos dados que exige a captura de uma quantidade enorme de variáveis do negócio antes não tratadas.

#### 5.8.3 Suporte a um Número Maior de Usuários

No passado, o processo de decisão estava concentrado nos níveis mais elevados da organização. Nos últimos anos, inúmeros fatores alteraram a estrutura das organizações e, consequentemente, os seus processos de decisão. Se antes as decisões eram tomadas pelo alto escalão, hoje em dia, existem muitas decisões que são tomadas pela gerência média e, em alguns casos, até pela linha de frente. Consequentemente, a necessidade por acesso a dados que permitam os colaboradores embasarem as suas decisões cresceu enormemente.

#### 5.8.4 O Ciclo de Vida e o Armazenamento dos Dados

À medida que o perfil dos dados vem mudando nos últimos anos, cresce a necessidade pela adoção de metodologias para a gestão do seu ciclo de vida. Mesmo com os avanços tecnológicos ainda é impossível, seja do ponto de vista técnico, seja do ponto de vista econômico, administrar todo o volume de dados das grandes organizações em um único repositório central. O que se observa é uma tendência por se armazenar os dados em função de seu perfil de acesso. Essa tendência foi bem caracterizada por Inmon et al no que diz respeito ao ambiente de Data Warehouse em seu último trabalho onde propõe o **DW 2.0** (INMON; STRAUSS; NEUSHLOSS, 2008). Dois dos estudos de caso apresentados neste trabalho também ilustram como as organizações administram o ciclo de vida de seus dados. Do ponto de vista de integração de dados o Data Warehouse continua a ser a principal opção para o suporte das necessidades de análise de qualquer organização. O crescente volume de dados e usuários vem obrigando as organizações a reverem sua infraestrutura de Data Warehouse. Por outro lado, o avanço tecnológico em diversas frentes como capacidade de processamento, de armazenamento aliado a novos paradigmas de banco de dados como os columnstores e os bancos de dados paralelos criou uma variedade grande de soluções para que as empresas enfrentem os novos desafios do alto volume.

O principal fator de decisão na atualização da infraestrutura de *Data Warehouse* nas grandes organizações é a questão da escalabilidade de dados e de usuários. Uma tendência observada na maior parte das grandes organizações é o crescente investimento em arquiteturas baseadas em processamento paralelo maciço (ou MPP – *Massive Parallel Processing*). As soluções de mercado baseadas nesse tipo de tecnologia são comumente chamadas de EDW (ou *Enterprise Data Warehousing*) (KOBIELUS, 2009).

#### 5.8.5 Necessidades Específicas para a Monitoração de Processos

No processo de monitoração, o perfil de utilização dos dados não se assemelha ao perfil tradicional do *Data Warehouse*, ao contrário do que ocorre no processo de análise. Nesta etapa, o objetivo principal é alertar ou permitir que os colaboradores possam identificar eventos críticos a ponto de tomarem as devidas ações. Segundo o modelo proposto por Hackathorn (ver seção 3.2) o que se deseja nessa etapa é reduzir a latência de análise de tal forma que as ações corretivas possam ser tomadas pelo pessoal da linha de frente.

Dependendo da dinâmica do ambiente de negócios onde a organização está inserida, o tempo hábil para a tomada de ações pode ser bastante curto, geralmente da ordem de alguns minutos e, consequentemente, a arquitetura de sistemas utilizada deve ser capaz de notificar os colaboradores de forma efetiva. Neste sentido, as ferramentas mais utilizadas são os *dashboards* ou painéis indicadores. Esses painéis são comumente implementados através de uma interface gráfica, através de um *web-browser* onde são apresentados os principais indicadores e métricas dos processos sendo monitorados. Existem uma série de elementos visuais tais como *gauges* (ou odômetros), gráficos, semáforos, indicadores estilo *Kanban* entre outros que facilitam o alerta visual de condições específicas.

Com relação ao perfil de acesso aos dados no processo de monitoração, o mais comum é acessar os mais recentes. Ao contrário do processo de análise, onde muitas vezes é preciso manipular dados dos últimos anos, no caso da monitoração,

os dados dos últimos dias ou de uma dezena de dias já são suficientes. Outro aspecto importante do perfil de acesso é que, enquanto nos processos analíticos boa parte das decisões é realizada com dados sumarizados, no processo de monitoração as decisões são realizadas comparando-se os valores correntes de alguns indicadores contra os valores sumarizados dos mesmos indicadores em períodos recentes ou contra os indicadores históricos. A título de exemplo, no caso de um indicador, como as vendas de um determinado produto em um ponto de venda, pode-se comparar o volume de vendas atual contra o volume de vendas no último dia ou nos últimos dias contra o volume de vendas no mesmo dia seis meses antes ou um ano antes.

Dentre as capacidades necessárias a uma arquitetura de sistemas para suportar tais requisitos tem-se:

- Armazenamento eficiente dos dados recentes;
- Possibilidade de sumarização dos dados recentes quase que instantaneamente;
- Possibilidade de armazenamento de valores de referência, ou seja, do histórico de indicadores;
- Definição de indicadores e de regras de avaliação dos mesmos.

Assim como a latência nos dados tem um papel importante no processo de aquisição dos dados ela também tem um papel importante na etapa de monitoração. Neste caso, o que se pretende é reduzir a latência de análise, o que implica na capacidade de sumarização dos dados recentes de acordo com a latência máxima tolerada. Do ponto de vista da integração dos dados em tempo real, até que esses dados sejam apresentados graficamente, há duas questões importantes que podem afetar a arquitetura de sistemas:

- O armazenamento dos dados integrados recentemente;
- A possibilidade de sumarização de dados correntes em tempo real, de acordo com os requisitos de latência de análise impostos pelo negócio.

#### 5.8.6 O Armazenamento dos Dados Integrados Recentemente

O ODS constitui a principal solução para o armazenamento de dados integrados durante a fase de monitoração. Com o avanço na capacidade de processamento, memória e armazenamento dos servidores atuais aliados ao desempenho oferecido pelos principais gerenciadores de bancos de dados relacionais é possível construir ambientes de ODS que suportem grandes volumes de dados e usuários e com alta frequência de atualização. Uma alternativa ao ODS, cuja adoção vem crescendo nos últimos anos, é a utilização dos *In-Memory Databases* para a integração e armazenamento dos dados recentes. Tais ferramentas permitem armazenar um volume de dados razoável, da ordem de dezenas de *gigabytes*, além de permitirem o processamento de consultas (comparações, sumarizações, etc) em tempo instantâneo. Uma tecnologia que também pode ser empregada neste sentido é a integração de informação (ou EII) em cenários específicos, onde existe a necessidade de acesso a fontes de dados heterogêneas.

Apesar da possibilidade de emprego de outras abordagens, o ODS deve permanecer como a principal opção para o armazenamento de dados integrados recentemente. Do ponto de vista do ciclo de vida dos dados, o ODS tem apresentado um papel cada vez mais importante atuando como caminho intermediário dos dados entre a sua captura e o seu posterior armazenamento no Data Warehouse para análise. Essa tendência pode ser observada a partir dos estudos de caso apresentados anteriormente (Quicken Loans e Yahoo!), assim como, a partir do conceito de DW 2.0 apresentado por Inmon et al recentemente.

## 5.8.7 A Frequência de Atualização e a Sumarização dos Dados em Tempo Real

A sumarização dos dados em tempo real pode representar o principal desafio em algumas implantações. Com o desenvolvimento tecnológico recente tanto em

hardware quanto em software é possível, nos dias de hoje, construir ambientes de ODS que suportem alguns terabytes de dados sendo acessados por centenas de usuários e com períodos de atualização bastante curtos. Nestas condições, nos dias de hoje, é perfeitamente possível construir soluções cuja frequência de atualização seja da ordem de dezenas de minutos. Ocorre que, atualmente, existem cenários onde é necessário trabalhar com frequência de atualização maiores, ou seja, com latências da ordem de poucos minutos ou de segundos. Nestes casos, tem-se a situação onde é necessária a adoção de uma nova alternativa tecnológica, como as plataformas de stream processing (STONEBRACKER; ÇETINTEMEL, 2005). Nestes casos, devido aos requisitos extremos para a latência de análise é preciso tratar a atualização dos dados como sendo feita a partir de um fluxo de dados.

A escolha das tecnologias de armazenamento e processamento de dados pode variar em função da latência de análise necessária. Na maior parte dos casos, os requisitos de latência de análise ainda são da ordem de dezenas de minutos. Nestes casos, a adoção do ODS ainda constitui a melhor alternativa do ponto de vista de integração dos dados. Existem duas outras tecnologias que podem ser somadas ao ODS com o objetivo de flexibilizar ou viabilizar a implantação de uma camada de integração virtual no curto-médio prazo. Essas tecnologias são os *In-Memory Databases* e as ferramentas de EII.

Em ambientes de negócio muito dinâmicos, onde a latência da ação precisa ser muito pequena e, consequentemente, a latência de análise também precisa seguir na mesma direção, faz-se necessário considerar a adoção de uma plataforma de stream processing. Um aspecto importante é que, mesmo nesses casos, a adoção do ODS ainda se faça necessária para o armazenamento de dados recentes.

# 5.8.8 O Uso de Avançadas Técnicas de Data Mining e Knowledge Discovery

A aplicação das técnicas de data mining e knowledge discovery vem alterando significativamente a forma como muitas organizações funcionam. As aplicações superam os processos convencionais de planejamento e operação também sendo utilizadas no projeto de novos produtos, na fidelização de clientes, no controle de fraudes e no gerenciamento do nível de satisfação dos funcionários. Essa mudança implica na captura, armazenamento e processamento de um volume maior de dados de duas formas: através de uma maior diversidade de assuntos tratados e uma maior granulosidade nos dados. A inteligência artificial vem sendo aplicada aos negócios desde a década de mil novecentos e oitenta. Mais recentemente, com o avanço tecnológico, o seu emprego tanto nos processos de análise como de execução de processos vem crescendo. Alguns autores como Davenport e Harris (2005; 2006) definem essas novas aplicações como sistemas de automatização de decisão. Tais sistemas, geralmente, são baseados em regras e, comumente, oferecem uma solução para uma área funcional da empresa (como finanças ou vendas) para um problema específico e repetitivo de negócios (TURBAN et al, 2008).

A automatização do processo decisório ou, conforme o modelo de maturidade proposto (seção 3.6), a execução de processos são, geralmente, alcançados através do agrupamento e compilação das experiências de usuários em um conjunto de regras de negócios que são incorporadas a um mecanismo de fluxo de trabalho. Tais sistemas trabalham a partir de um conjunto de eventos de negócios que são submetidos às regras programadas no sistema e, por final, uma ou mais ações são tomadas automaticamente.

Segundo Davenport e Harris (2005) existem muitas aplicações bem sucedidas de sistemas de automatização de decisão como:

- Configuração de produtos e serviços;
- Otimização de preços;
- Decisões de encaminhamento ou segmentação;
- Auditoria contínua (verificação de conformidades);
- Detecção de fraudes;
- Previsão dinâmica de vendas.

### 5.9 Considerações

A adoção da Inteligência de Negócios em Tempo Real requer investimentos consideráveis. As organizações, buscando sua adoção, devem levar em consideração as limitações do ambiente de *Data Warehouse* e adaptá-lo através da adoção gradual de mecanismos de integração para a redução das latências, principalmente a latência de dados. Como normalmente uma pequena parcela dos processos de negócio tem potencial de ganho com o uso de dados integrados em tempo real, recomenda-se a priorização dos investimentos segundo o modelo de avaliação de McGee. A seleção dos mecanismos de integração deve levar em conta o ciclo de vida dos dados, a variedade de assuntos e fontes, os volumes de dados e de usuários, além dos processos de captura, atualização e sumarização de dados.

#### 6 RELATO DE CASO

Os conceitos apresentados no capítulo anterior foram aplicados na proposição de uma solução de *Real-Time Business Intelligence* para uma empresa do setor financeiro. A opção pela aplicação em um caso real se justifica pela oportunidade de por em prática a teoria. O objetivo desse estudo foi apresentar um exemplo prático de como as novos mecanismos de integração de dados, em particular o *Stream Processing*, podem viabilizar a redução de latência nos dados a intervalos muito curtos.

O caso proposto refere-se à implantação de uma prova de conceito onde o principal objetivo era demonstrar que a proposta de solução apresentada atendia aos objetivos de negócio e aos principais indicadores de latência de dados.

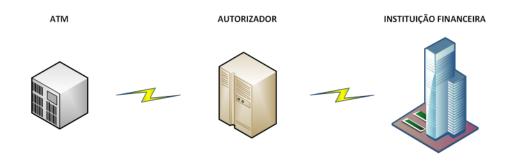


FIGURA 22 - Rede de ATMs

A empresa em questão utiliza uma rede de caixas automáticos, também conhecidos como ATM, ou *Automated Teller Machine*. Essa empresa possui milhares de caixas automáticos espalhados pelo território nacional realizando mais de um milhão de transações por mês. A qualidade e efetividade do serviço prestado depende diretamente da quantidade de transações efetuadas com sucesso.

Há alguns anos, essa empresa investiu em sua infraestrutura de *Business Intelligence* ao criar um *Data Warehouse* para consolidar as informações da rede de ATMs. Os relatórios gerencias para as equipes responsáveis pela gestão da rede são gerados a partir desse *Data Warehouse*. Além disso, o *Data Warehouse* serve

como fonte para a alimentação de uma dezena de *dashboards* que foram desenvolvidos e são utilizados pelos mais diferentes níveis da empresa, desde os diretores até os funcionários responsáveis para operação e manutenção da rede.

### 6.1 O Sistema BAM (Business Activity Monitoring)

A infraestrutura de monitoração de operações dessa empresa é composta por um sistema denominado internamente de sistema BAM, ou *Business Activity Monitoring*. Este sistema por sua vez, compreende um *Data Warehouse* relacional, uma *Staging Area* e um *Data Mart* multidimensional como se vê na figura 23.

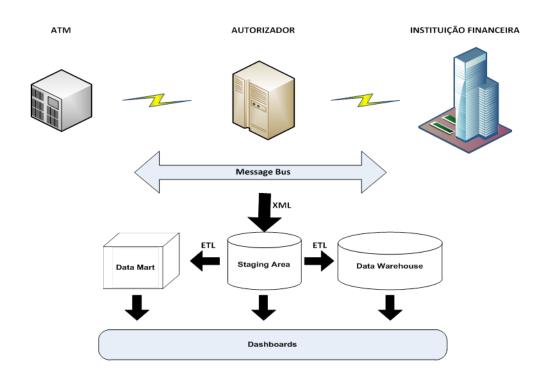


FIGURA 23 - Ambiente atual de Business Activity Monitoring

Toda nova transação, seja ela bem ou mal sucedida, gera uma mensagem XML em um *Message Bus* que, por sua vez, a distribui, por meio de filas, para diversas aplicações. A cada um ou dois minutos essas mensagens são carregadas, através de um processo de *Trickle-Feed*, em uma base de dados que atua como uma

Staging Area. O processo de carga, nesta etapa, é o responsável por transformar os dados de modo que estes sejam armazenados em tabelas nessa base de dados.

Os dados carregados nessa *Staging Area* são transformados e carregados, de hora em hora, no *Data Warehouse*. O *Data Warehouse* constitui a principal fonte para a geração de informações gerenciais oferecendo uma visão do negócio com uma defasagem ou latência nos dados de uma hora. Os mesmos dados são transformados e carregados em um *Data Mart* operacional que serve como principal fonte de indicadores operacionais que são apresentados nas dezenas de *dashboards* utilizados para monitorar o negócio. No início do projeto, pretendia-se atualizar esses dados com uma latência máxima de cinco minutos mas, devido a problemas com o desempenho desse sistema, a empresa viu-se obrigada a trabalhar com uma latência de quinze a vinte minutos, dependendo do volume de transações a ser carregado de uma só vez. Dentre os principais indicadores de atividade registrados por esse sistema, pode-se citar:

- 1) Total de transações em um dado instante;
- 2) Total de transações por tipo de transação, em transações por minuto, em um dado instante:
- Percentual de ATMs sem transações, em percentuais;
- 4) Quantidade de ocorrências (incidentes), no dia corrente;
- 5) Volume de transações corrente, por tipo de transação, comparado a patamares históricos.

#### 6.2 O Problema da Latência nos Dados

Neste caso, o problema da latência nos dados transformou-se em um grande desafio para o negócio. À medida que o negócio cresce, seja pelo acréscimo de

mais ATM's, seja pelo volume total de transações em potencial, maior a necessidade pela obtenção de informações em tempo real que permite a empresa gerir de forma eficiente a rede.

Com base nas perspectivas de crescimento de ATM's e do volume de operações, podendo atingir até 400.000 operações por hora, a direção da empresa decidiu que era preciso buscar outra plataforma de BAM, em substituição à plataforma existente, que permitisse suportar o volume de transações de aproximadamente 600 transações por segundo e, ao mesmo tempo, oferecer dados operacionais com uma latência de no máximo 1 minuto.

Para a empresa, é vital manter a rede operando o tempo todo. Cada tipo de incidente, seja a indisponibilidade parcial ou total de um ATM prejudica o resultado da empresa. Existem muitos cenários de falha que podem provocar prejuízos. Os cenários de maior impacto são aqueles onde uma parcela significativa dos clientes ficam sem o serviço. Isto pode ser ocasionado por uma falha na rede ou por uma falha em sistemas internos. À medida que o volume de transações cresce, aumenta também a complexidade de gerir a rede e o potencial para a ocorrência de incidentes. Por este motivo, a empresa precisa estar preparada para monitorar a operação de forma efetiva.

# 6.3 Requisitos para a Latência nos Dados

Os requisitos para a redução na latência que foram colocados pela empresa, tendo em vista as projeções de crescimento do negócio, são:

- Volume de transações em período de pico em 600 transações por segundo;
- Latência nos dados correntes de até 1 minuto;
- Comparação de dados correntes com patamares nos últimos quinze e trinta minutos;
- Volume diário de transações de três a cinco milhões de operações.

## 6.4 Alternativas para a Redução das Latências

A partir da análise de cada um dos elementos do sistema *BAM* existente, foi possível avaliar se o sistema poderia ser adaptado de modo a atender aos requisitos ou se precisaria de alguma substituição.

Com relação ao *Data Warehouse*, este cumpre o seu papel de ser a fonte para a geração de relatórios gerenciais. Como não existe uma necessidade de redução na latência atual, na ordem de uma hora, esse componente pode ser mantido da forma relatada. Como, neste caso, o *Data Warehouse* não serve a objetivos operacionais, não se observam os problemas destacados na seção 5.4.2.

Entretanto, com relação do *Data Mart* multidimensional, todos esses problemas podem ser observados por seu enfoque ser operacional. Neste caso, os problemas destacados por Stonebracker *et al* para o modelo de processamento *"inbound"* se acentuam pois, além da captura e armazenamento dos dados, como se trata de uma base de dados multidimensional é preciso que boa parte das agregações sejam précalculadas, o que por sua vez introduz mais latência nos dados. A substituição desse *Data Mart* da tecnologia multidimensional pela relacional também não resolve o problema devido aos fatores expostos na seção 5.4.2. A única solução, como exposto na seção 5.5, é buscar tecnologias alternativas, que permitam reduzir a latência nos dados com base nos requisitos colocados. Em função da latência para os dados correntes, resta apenas explorar o *Stream Processing*, uma vez que, o modelo de processamento *"inbound"* não permite a atualização e realização de consultas em intervalos de tempo tão pequenos (STONEBRAKER; ÇETINTEMEL, 2005).

Com relação à *Staging Area*, esta, aparentemente, atua apenas como um elemento retardador, acrescentando mais latência ao processo. Este componente pode ser substituído por um *ODS*, sendo alimentado por um processo de *Trickle-Feed* a partir do *Message Bus*, em intervalos bastante curtos, como a cada um ou dois minutos. Esse *ODS* pode armazenar os dados do dia corrente e servir como fonte de dados para a atualização do *Data Warehouse*. Ele também serve como fonte de dados

para o cálculo dos patamares operacionais (valores médios dos indicadores nos últimos quinze ou trinta minutos).

Uma possibilidade existente é substituir tanto a *Staging Area* quanto o *ODS* pelo *Stream Processing*. Foram identificadas duas alternativas viáveis neste caso.

### A figura 24 ilustra a primeira alternativa identificada:

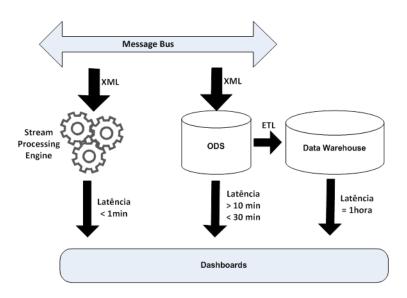


FIGURA 24 - Primeira alternativa (uso de *Stream Processing* e ODS)

Neste caso, vale ressaltar que o *Stream Processing Engine* cumpre tanto o papel de aquisição quanto o de armazenamento temporário dos dados. Neste cenário, esta tecnologia é utilizada apenas para a entrega dos dados integrados (com transformações e agregações) e correntes, ou seja, que foram capturados nos últimos 60 segundos. Do ponto de vista de armazenamento dos dados, tem-se o *ODS* para dados com latência mínima de dez minutos. Os dados são armazenados no *ODS* por um intervalo não superior a três dias, sendo que idealmente, depois de carregados no *Data Warehouse*, os dados presentes no *ODS* podem ser descartados. A quadro 10 ilustra o ciclo de vida dos dados nesse cenário:

Elemento	Latência nos Dados	Período de	
	(máxima)	Retenção/Armazenamento	
		dos dados	
Stream Processing Engine	1 minuto	3-5 minutos	
Operational Data Store	30 minutos	1-3 dias	
Data Warehouse	1 hora	3 anos	

QUADRO 9 - Latência e Retenção na primeira alternativa

### A figura 25 ilustra a segunda alternativa identificada:

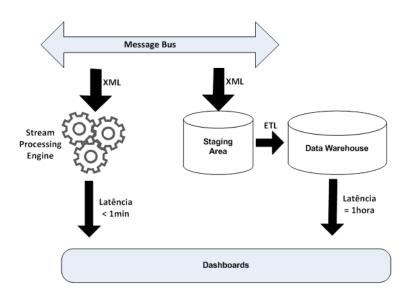


FIGURA 25 - Segunda alternativa (uso exclusivo do Stream Processing, sem o uso do ODS)

Neste caso, o *Stream Processing Engine* cumpre tanto o papel de aquisição quanto de armazenamento temporário dos dados, porém com uma retenção maior dos dados (por no mínimo 30 minutos). O ODS é totalmente substituído pelo *Stream Processing Engine* que assume o seu papel de entregar os indicadores agregados nos últimos quinze e trinta minutos. A *Staging Area* cumpre dois propósitos: alimentar o *Data Warehouse* e ao mesmo tempo retirar dele a sobrecarga gerada pela inserção constante de dados que é provocada pelo processo de aquisição de dados utilizando-se a técnica de *Trickle-Feed*. O quadro 11 ilustra o ciclo de vida dos dados nesse cenário:

Elemento	Latência nos Dados	Período de
	(máxima)	Retenção/Armazenamento
		dos dados
Stream Processing Engine	1 minuto	30 minutos
Data Warehouse	1 hora	3 anos

QUADRO 10 - Latência e Retenção na segunda alternativa

No caso em questão, as duas alternativas atendem aos requisitos colocados. O quadro 12 apresenta as vantagens e desvantagens de cada abordagem:

Alternativa	Vantagens	Desvantagens
1 <sup>a</sup>	O ODS atua como uma camada	O desempenho pode ser um
	de persistência para os	pouco inferior à outra alternativa,
	patamares, ou seja, indicadores	uma vez que, é necessário
	avaliados nos últimos quinze e	integrar os dados do ODS
	trinta minutos. Esses dados	(patamares) no Stream
	podem ser facilmente recuperados	Processing Engine.
	a qualquer momento.	
2ª	Essa abordagem tende a	Não existe uma persistência dos
	apresentar um melhor	patamares uma vez que os
	desempenho uma vez que a maior	Stream Processing Engines
	parte dos dados utilizados nos	operam em memória RAM. Em
	dashboards operacionais	caso de falha ou reinicialização
	encontram-se no próprio Stream	do serviço é preciso aguardar
	Processing Engine.	por até trinta minutos para que
		seja possível aferir os novos
		patamares. Uma alternativa é
		consultar a Staging Area para
		gerar a mesma informação.

QUADRO 11 - Comparativo entre alternativas

#### 6.5 Ciclo de Vida dos Dados

Assim como foi proposto por Inmon no *DW 2.0* (INMON; STRAUSS; NEUSHLOSS, 2008) e pode ser observado em outros casos reais de aplicação, tal como apresentado nas seções 4.4.1 e 4.4.2, neste relato de caso, os dados fluem em diferentes setores de armazenamento e processamento, que por sua vez podem fazer uso de diferentes tecnologias.

#### 6.6 Prova de Conceito

Durante as discussões sobre as duas alternativas propostas optou-se pela realização de uma prova de conceito para a avaliação da tecnologia de *Stream Processing*. A equipe responsável pela avaliação considerou desnecessária a avaliação do *ODS*, uma vez que se trata de um conceito consolidado. Neste caso, optou-se por uma solução similar à segunda alternativa proposta anteriormente, ou seja, o *Stream Processing Engine* ficou responsável tanto por consolidar os dados correntes como pelos patamares. A figura 26 ilustra o cenário de teste utilizado nesta prova de conceito.

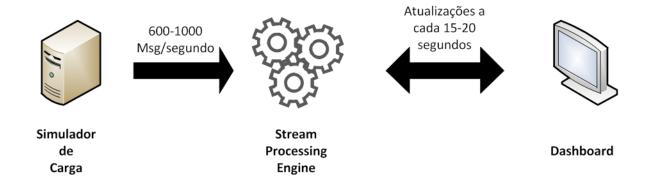


FIGURA 26 - Prova de Conceito

Para a realização dos testes, utilizou-se uma aplicação desenvolvida especialmente para essa prova de conceito para atuar como um simulador de carga, responsável

por gerar mensagens XML. A partir dos equipamentos utilizados, foi possível atingir uma taxa de aproximadamente mil mensagens por segundo. A plataforma de *Stream Processing Engine* utilizada comportou-se bem, respondendo aos requisitos de latência colocados, uma vez que, em outros testes realizados, foi capaz de processar até 50 mil mensagens por segundo. Outra particularidade dessa prova de conceito foi a substituição da interface tradicional dos *dashboards* por uma interface em estilo *web application*, baseada em *plug-ins* de interface rica (ou *Rich Internet Applications*) que permitiram fazer atualizações periódicas nos dados a cada 15-20 segundos.

#### 6.7 Resultados Obtidos

Os resultados foram satisfatórios na medida em que os requisitos de latência de dados foram atendidos e a tecnologia de *Stream Processing* além de se tratar da única opção viável tecnicamente, mostrou-se bastante estável e robusta durante a realização dos testes. A prova de conceito serviu para mostrar que a empresa precisa investir em novas tecnologias para reduzir sua latência nos dados e possibilitar o uso eficiente dos *dashboards*. A versão anterior do sistema baseia-se em ferramentas tradicionais para a geração de *dashboards* e que também não se comportam muito bem com a atualização de dados em intervalos tão pequenos. As limitações dessa versão estão relacionadas tanto à capacidade de se trabalhar com baixa latência nos dados quanto na capacidade em atualizar as informações em sua interface gráfica. A tecnologia de interface rica, adotada na prova de conceito, mostrou ser uma opção viável e de menor esforço de desenvolvimento.

## 7 CONCLUSÕES

A integração dos dados tem um papel fundamental na redução das latências nos dados e, consequentemente, na adoção da inteligência empresarial em tempo real. Uma empresa interessada em adquirir a competência de prever e se antecipar aos eventos relevantes para o seu negócio precisa planejar de forma eficiente a atualização de sua infraestrutura tecnológica para tirar proveito das informações em tempo real. Embora o *Data Warehouse* seja o principal mecanismo utilizado para a integração e distribuição de dados nas empresas, ele apresenta uma série de limitações quando é necessário trabalhar com dados atualizados muito frequentemente (baixa latência nos dados). Nessas situações, faz-se necessário a adoção de tecnologias complementares para a integração de dados e distribuição dos dados em tempo real.

#### 7.1 RESULTADOS OBTIDOS

A principal contribuição desta dissertação foi apresentar as tendência por um enfoque mais operacional na inteligência de negócios, as limitações do ambiente tradicional de *Data Warehouse* em atendê-las e destacar a importância que a adoção de novos mecanismos de integração de dados em tempo real têm na efetiva implantação da inteligência de negócios em tempo real.

Outras contribuições que podem ser mencionadas são:

- Uma revisão da literatura prática e acadêmica, compilando as principais abordagens para a questão da inteligência empresarial em tempo real;
- Apresentação das tendências em inteligência empresarial e a importência da redução na latência dos dados;
- Apresentação das limitações do Data Warehouse como principal mecanismo para a integração dos dados em tempo real;

- Indicar alternativas para a atualização do ambiente de Data Warehouse com a adoção de tecnologias complementares de integração de dados, com o objetivo de permitir uma redução na latência nos dados;
- Comparar as diferentes tecnologias de integração de dados de modo a estabelecer critérios para a seleção diante de requisitos de latência de dados.

### 7.2 TRABALHOS FUTUROS

Como oportunidades para a realização de trabalhos futuros, podem ser desenvolvidos estudos teóricos e práticos relacionados tanto ao desenvolvimento das tecnologias de integração quanto à sua aplicação em casos reais.

Nos aspectos teóricos, a demanda pelo desenvolvimento de mecanismos mais eficientes para a redução de latência tem motivado a realização de muitas pesquisas nas áreas de bancos de dados e integração de sistemas e aplicações.

Com relação aos aspectos mais práticos, a busca por casos reais de aplicação tem recebido bastante atenção de publicações com orientação comercial. Infelizmente, ainda se observa uma menor atenção de publicações de caráter acadêmico.

# 8 REFERÊNCIAS BIBLIOGRÁFICAS

AGOSTA, Lou. **Data Warehousing Refresh Rates**. Information Management Magazine, June 2003. Disponível em: <a href="http://www.information-management.com/issues/20030601/6834-1.html">http://www.information-management.com/issues/20030601/6834-1.html</a> . Disponível para Download em: Acesso em 01 fev. 2011.

\_\_\_\_\_. **Is Your Data Warehouse Active ?** Information Management Magazine, April 2004.

AHUJA, Mona; CHEN, Cheng Che; GOTTAPU, Ravi; HALLMANN, Jörg; HASAN, Waqar; JOHNSON, Richard; KOZYRCZAK, Maciek; PABBATI, Ramesh; PANDIT, Neeta; POKURI, Sreenivasulu; UPPALA, Krishna. **Peta-Scale Data Warehousing at Yahoo!** SIGMOD'09, June 29-July 2, 2009, Providence, Rhode Island, USA.

ANDERSON-LEHMAN, Ron; WATSON, Hugh J.; WIXOM, Barbara H.; HOFFER, Jeffrey A. Continental Airlines Fly High with Real-Time Business Intelligence. MIS Quarterly Executive, Vol. 3, No. 4, December 2004.

AXSON, David A. Best Practives in Planning and Performance Management: From Data to Decisions. John Wiley & Sons, 2007.

AZIZA, Bruno; FITTS, Joey. **Drive Business Performance: Enabling a Culture of Intelligent Execution.** John Wiley & Sons, 2008.

AZVINE, B.; CUI, Z.; NAUCK, D.D.; MAJEED, B. Real Time Business Intelligence for the Adaptive Enterprise. Proceedings of the 8th IEEE International Conference on E-Commerce Technology and the 3rd IEEE International Conference on Enterprise Computing, E-Commerce, and e-Services 2006 (CEC/EEE'06).

BERNSTEIN, Philip A.; HASS, Laura M. Information Integration in the Enterprise. Communications of the ACM, Vol. 51, Nro. 9. September, 2008.

BUYTENDIJK, F.; WOOD, B.; GEISHECKER, L. **Drivers and Challenges of Corporate Performance Management.** Gartner Strategic Analysis Report R-22-0730, January 2004.

BRUCKNER, Robert M.; LIST, Beate; SCHIEFER, Josef. **Striving towards Near Real-Time Data Integration for Data Warehouses.** Proceedings of the 4th International Conference on Data Warehousing and Knowledge Discovery. LNCS 2454, Springer-Verlag, 2002, pp 317-326.

CASTELLANOS, Malu; SIMITSIS, Alkis; WILKINSON, Kevin; DAYAL, Umeshwar. Automating the Loading of Business Process Data Warehousing. 12th International Conference on Extending Database Technology, Saint Pertersburg, Russia, March 24-26, 2009. ACM International Conference Proceeding Series.

Data Integration Flows for Business Intelligence.	12th International
Conference on Extending Database Technology, Saint Pertersbu	ırg, Russia, March
24-26, 2009. ACM International Conference Proceeding Series.	

\_\_\_\_\_. Business Processes Meet Operational Intelligence. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering (2009).

CLARK, Thomas D.; JONES, Mary C.; ARMSTRONG, Curtis P. **The Dynamic Structure of Management Support Systems: Theory Development, Research Focus, and Direction.** MIS Quarterly Vol. 31, No 3, pp. 579-615, September 2007.

CONN, Samuel S. OLTP and OLAP Data Integration: A Review of Feasible Implementation Methods and Architectures for Real Time Data Analysis. Proceedings of the IEEE Southeastcon 2005.

CORCORAN, Michael. **Six BI Trends You Can't Ignore.** TWDI Research. July 23, 2008. Disponível em: <a href="http://esj.com/articles/2008/07/23/six-bi-trends-you-cant-ignore.aspx">http://esj.com/articles/2008/07/23/six-bi-trends-you-cant-ignore.aspx</a> . Acesso em 01 fev. 2011.

DAVENPORT, Thomas H.; HARRIS, Jeanne G. Automated Decision Making
Comes of Age. Management of Information Systems, Operations Management and
Research. MIT Sloan Management Review, July 2005.
Competing on Analytics. Harvard Business Review. (January 2006).
DAYAL, Umeshwar; KUNO, Harumi; WIENER, Janet L.; GANAPATHI, Archana;
KROMPASS, Stefan. Managing Operational Business Intelligence Workloads.
ACM SIGOPS Operating Systems Review, Volume 43, Issue 1, January 2009.
DRESDNER, Howard. Business Activity Monitoring: 'New Age' BI ? Gartner Inc.
Research Note LE-15-8377, 2002.
ECKERSON, Wayne W. The Soft-Side of Real-Time BI. DM Review, August 2004.
Best Practices in Business Performance Management and Technical
Strategies. TWDI Report, The Data Warehouse Institute, 2004. Disponível em:
http://download.101com.com/tdwi/research_report/2004_Best_Practices_Business_
Report.pdf . Acesso em 01 fev. 2011.
Performance Dashboards: Measuring, Monitoring and Managing Your
Business. John Wiley & Sons, 2006.
Best Practices in Operational BI – Converging Analytical and
Operational Processes. TWDI Report, The Data Warehouse Institute, 2007.
Operational BI: Sorting Out Your Options. TWDI Research. January 14,
2010. Disponível em: <a href="http://tdwi.org/articles/2010/01/14/operational-bi-options.aspx">http://tdwi.org/articles/2010/01/14/operational-bi-options.aspx</a>
. Acesso em 01 fev. 2011.
FEW Clauban Bookhoond Confusion Intelligent Entermine Managine Managh 00
FEW, Stephen. <b>Dashboard Confusion</b> . Intelligent Enterprise Magazine. March, 20,
2004. Disponível em: <a href="http://intelligent-">http://intelligent-</a>
enterprise.informationweek.com/showArticle.jhtml;jsessionid=J4E55D4UCEI0TQE1G
HRSKHWATMY32JVN?articleID=18300136&pgno=1 . Acesso em 01 fev. 2011.

FEW, Stephen. Information Dashboard Design. O'Reilly & Associates Press, 2006.

GOLFARELLI, Matteo; RIZZI, Stefano; CELLA, Iuris; **Beyond Data Warehousing: What's Next in Business Intelligence**. DOLAP'04, Association of Computing Machinery, 2004.

GOVEKAR, Milind; SCHULTE, Roy. **BAM Architecture: More Building Blocks Than You Think.** Gartner Inc. Research Note AV-15-5070.2002.

GREINER, Torsten; DÜSTER, Willy; POUATCHA, Francis; VON AMMON, Rainer; BRANDL, Hans-Martin; GUSCHAKOWSKI, David. Business Activity Monitoring of Norisbank taking the example of the application easyCredit and the future adoption of Complex Event Processing (CEP). Proceedings of the 4th International Symposium on Principles and Practice of Programming in Java. Association for Computing Machinery, 2006.

HACKATHORN, Richard. **Current Practices in Active Data Warehousing**. DM Review White Paper, November 2002.

HACKATHORN, Richard. **The BI Watch: Real-Time to Real-Value.** DMReview, January 2004. Disponível em: <a href="http://www.bolder.com/pubs/DMR200401-Real-Time%20to%20Real-Value.pdf">http://www.bolder.com/pubs/DMR200401-Real-Time%20to%20Real-Value.pdf</a> . Acesso em 01 fev. 2011.

HALEVY, Alon Y.; ASHISH, Naveen; BITTON, Dina; CAREY, Michael; DRAPER, Denise; POLLOCK, Jeff; ROSENTHAL, Arnon; SIKKA, Vishal. **Enterprise Information Integration: Successes, Challenges and Controversies.** SIGACM-SIGMOD 2005. Baltimore, USA.

HOUGHTON, Robert; EL SAWY, Omar A.; GRAY, Paul; DONEGAN. Craig; JOSHI, Ashish. Vigilant Information Systems for Managing Enterprises in Dynamic Supply Chains: Real-Time Dashboards at Western Digital. MIS Quarterly Executive Vol. 3 No 1 / March 2004.

HOWSON, Cindy. **Take Advantage of In-Memory Analytics**. Intelligent Enterprise Magazine. April 20, 2009. Disponível em: <a href="http://intelligent-enterprise.informationweek.com/channels/business\_intelligence/showArticle.jhtml;jsessionid=XJXGGDGMGILPZQE1GHPCKHWATMY32JVN?articleID=216900096">http://intelligent-enterprise.informationweek.com/channels/business\_intelligence/showArticle.jhtml;jsessionid=XJXGGDGMGILPZQE1GHPCKHWATMY32JVN?articleID=216900096</a>. Acesso em 01 fev. 2011.

INMON, Bill. **Designing the Operational Data Store**. Information Management Magazine, July 1998. Disponível em: <a href="http://www.information-management.com/issues/19980701/469-1.html">http://www.information-management.com/issues/19980701/469-1.html</a>. Acesso em 01 fev. 2011.

INMON, Bill. **ODS Types / Information Management: Charting the Course.** Information Management Magazine, January 2000. Disponível em: <a href="http://www.information-management.com/issues/20000101/1749-1.html">http://www.information-management.com/issues/20000101/1749-1.html</a>. Acesso em 01 fev. 2011.

INMON, Bill; Strauss, Derek; Neushloss, Genia. **DW 2.0: The Architecture for the Next Generation Data Warehousing**. Morgan Kaufmann, 2008.

ITALIANO, Isabel C.; FERREIRA, João E. Synchronization Options for Data Warehouse Designs. IEEE Computer Society, March 2006.

KAPLAN, Robert S.; NORTON, David P. **The Balanced Scorecard – Measusres that Drive Performance.** Harvard Business Review, January-Febrary 1992.

KARAKASIDIS, Alexandros; VASSILIADIS, Panos; PITOURA, Evaggelia. **ETL Queues for Active Data Warehousing.** Second International ACM SIGMOD Workshop on Information Quality in Information Systems, Baltimore, MD, US, June 2005.

KOBIELUS, James. The Forrester Wave: Enterprise Data Warehousing Platforms, Q1 2009. Forrester Research, 2009.

KÜNG, Peter; HAGEN, Claus; RODEL, Marisa; SEIFERT, Sandra. Business Process Monitoring & Measurement in Large Bank: Challenges and selected

**approaches.** Proceedings of the 16h International Workshop on Database and Expert Systems Applications (DEXA'05). IEEE Computer Society, 2005.

LANGSETH, J.; VIVATRAT, N. **Outward Bound.** Intelligent Enterprise, November 2002.

LOFSTROM, Eric. Quicken Loans Relies on ProClarity to Help Analyze 3 Million Rows of Data Per Day. TDWI Research. Volume 17, May 2004. Disponível em: <a href="http://www.information-management.com/issues/20041001/1011114-1.html">http://www.information-management.com/issues/20041001/1011114-1.html</a>. Acesso em 01 fev. 2011.

LOSHIN, David. **Business Intelligence: The Savvy Manager's Guide**. Morgan Kaufmann Publishers, 2003.

LUBINSKI, Tom. Business Activity Monitoring: Process Control for the Enterprise. SL Corporation, 2008. Disponível em: <a href="http://www.sl.com/pdfs/BusinessActivityMonitoring-tom-lubinski.pdf">http://www.sl.com/pdfs/BusinessActivityMonitoring-tom-lubinski.pdf</a>. Acessado em: 01 fev. 2011

LUHN, Hans P. A Business Intelligence System. IBM Journal. October 1958.

LUCKHAM, David. <u>The Beginnings of IT Insight: Business Activity Monitoring</u>. EbizQ Magazine, June, 2004. Disponível em:

http://www.ebizq.net/topics/cep/features/4689.html .

Acesso em 01 fev. 2011.

MATIAS, Yossi. **Trends in High Performance Analytics.** SIGMOD 2006, June 27-29, 2006, Chicago, IL, USA.

MCCOY, David. **Business Activity Monitoring: Calm Before the Storm**. Gartner Inc. Research Note LE-15-9727, 2002.

MCGEE, Ken. **Gartner Updates Its Definition of Real-Time Enterprise.** Gartner Inc. Research Note DF-22-2973, 2004.

MCGEE, Ken, Heads Up: How to Antecipate Business Surprises and Seize Opportunities First. Harvard Business School Press 2004.

NESAMONEY, D. **BAM: Event-Driven Business Intelligence for the Real-Time Enterprise.** DM Review, March 2004.

PANIAM, Zeljko. **Just-in-Time Business Intelligence and Real-Time Decisioning**. International Journal of Applied Mathematics and Informatics. Issue 1, Volume 1, 2007.

POLITES, Greta L. From Real-Time BI to the Real-Time Enterprise: Organizational Enablers of Latency Reduction. Twenty-Seventh International Conference on Information Systems, Milwaukee 2006. Disponível em: http://aisel.aisnet.org/icis2006/85. Acesso em 01 fev. 2011.

POWER, Daniel J. **A Brief History of Decision Support Systems.** DSSResources.COM. Version 4.0, March 2007. Disponível em: http://dssresources.com/history/dsshistory.html. Acesso em 01 fev. 2011.

PUTTANGUNTA, Naveen. **Active Data Warehougins and the Quest for Real-Time BI.** InfoManagement Direct, October 2006.

REDDY, Ram. **The Reality of Real Time.** Intelligent Enterprise Magazine. June 2004. Disponível em:

SANTOS, Ricardo J.; BERNARDINO, Jorge. **Real-Time Data Warehouse Loading Methodology.** Twelfth International Database Engineering & Applications Symposium, Coimbra, Portugal, September 10-12, 2008.

SAWHNEY, M. **Net-Gains - Real-Time Reality Check.** CIO Magazine. March 2003. Disponível em: <a href="http://www.cio.com/archive/030103/netgains.html">http://www.cio.com/archive/030103/netgains.html</a>. Acesso em 01 fev. 2011.

SWOYER, Stephen. **When Right-Time Isn't Real-Time Enough**. TDWI Article. May 30, 2007. Disponível em: <a href="http://tdwi.org/articles/2007/05/30/when-righttime-isnt-realtime-enough.aspx?sc\_lang=en">http://tdwi.org/articles/2007/05/30/when-righttime-isnt-realtime-enough.aspx?sc\_lang=en</a> . Acesso em 01 nov. 2011.

STALK, George Jr. **Time – The Next Source of Competitive Advantage.** Harvard Business Review, July-August 1988.

STONEBRAKER, Michael; ÇETINTEMEL Uğur. "One Size Fits All": An Idea Whose Time Has Come and Gone. pp. 2-11, Proceedings of the 21st International Conference on Data Engineering, IEEE Computer Society Press, Tokyo, Japan, April 2005, ISBN 0-7695-2285-8.

STONEBRACKER, Michael et al. **Claremont Report on Database Research.** Communications of the ACM. June, 2009.

SWOYER, Stephen. When Right-Time Isn't Real-Time Enough. TDWI Article. May 30, 2007. Disponível em: <a href="http://tdwi.org/articles/2007/05/30/when-righttime-isnt-realtime-enough.aspx?sc\_lang=en">http://tdwi.org/articles/2007/05/30/when-righttime-isnt-realtime-enough.aspx?sc\_lang=en</a> . Acesso em 01 fev. 2011.

TAYLOR, Richard. **Concurrency in Data Warehouse.**Proceedings of the 26<sup>th</sup> International Conference on Very Large Databases, Cairo, Egypt, 2000

TURBAN, Efraim; SHARDA, Ramesh; ARONSON, Jay E.; KING, David. **Business Intelligence – Um Enfoque Gerencial para a Inteligência do Negócio**. Artmed Editora S.A, 2008.

WALLS, Joseph G.; WIDMEYER, George R.; EL SAWY, Omar A. **Building an Information System Design Theory for Vigilant EIS.** Information Systems Research, March 1992.

WATSON, Hugh J.; WIXOM, Barbara H. **The Current State of Business Intelligence.** IEEE Computer Society, Volume 40, Issue 9, September 2007.

WHITE, Colin. **Building the Real-Time Enterprise.** TDWI Report Series. November 2003.

WHITE, Colin. **Now is the Right Time for Real-Time BI.** Information Management Magazine. September 2004. Disponível em: <a href="http://www.information-management.com/issues/20040901/1009281-1.html?pg=1">http://www.information-management.com/issues/20040901/1009281-1.html?pg=1</a>. Acesso em 01 fev. 2011.

WHITEHORN, Mark. **Active Data Warehousing explained and examined.** SearchDataManagement.com, January 2008.