

DATA INTELLIGENCE

COLOCANDO INTELIGÊNCIA EM SEUS NEGÓCIOS



LISTA DE FIGURAS

Figura 1 – Google Trends BA, BI e EI	8
Figura 2 – Componentes do BI.....	10
Figura 3 – Processos do BI	14
Figura 4 – Perspectivas do BSC	16
Figura 5 – Modelo dimensional de venda.....	17
Figura 6 – Visão geral do BI	22

EXEMPLO

SUMÁRIO

1 COLOCANDO INTELIGÊNCIA EM SEUS NEGÓCIOS	4
1.1 O que é Business Intelligence?	4
1.2 Um breve histórico sobre BI	4
1.3 Definições atuais sobre BI	7
1.4 BI vs. BA.....	8
1.5 O conjunto que forma o BI.....	9
1.5.1 Dados do BI.....	10
1.5.1.1 Fontes OLTP	10
1.5.1.2 Fontes Externas	11
1.5.1.3 Staging Area (SA)	11
1.5.1.4 Operational Data Store (ODS).....	11
1.5.1.5 Data Warehouse (DW)	12
1.5.1.6 Data Marts (DM).....	12
2 PROCESSOS DO BI	14
2.1 Business Drivers	14
2.2 Modelagem de dados	17
2.3 ETL.....	18
2.4 Administração.....	19
2.5 Análise dos dados	19
2.6 Query and Reporting	20
2.7 OLAP (Processamento Analítico On-Line)	20
2.8 Data Mining	22
CONCLUSÃO.....	23
REFERÊNCIAS.....	24
GLOSSÁRIO	26

1 COLOCANDO INTELIGÊNCIA EM SEUS NEGÓCIOS

O que antes era um diferencial, hoje o uso da inteligência nos negócios (Business Intelligence) é um pilar para as tomadas de decisão de uma empresa.

Após criarmos sistemas e aplicativos, chegou a hora de planejar o futuro da sua startup com o uso do BI.

O processo de coleta, organização e, finalmente, análise de dados começa aqui!

1.1 O que é Business Intelligence?

Existem diversas respostas para essa pergunta, mas, com certeza, a definição atual mais conhecida para BI é a atribuída a Howard Dresner (posteriormente, pesquisador do Gartner Group). Dresner propôs, em meados de 1989, Business Intelligence como um termo guarda-chuva para descrever um conjunto de conceitos e métodos para melhorar a tomada das decisões de negócio, usando sistemas de suporte baseados em fatos.

1.2 Um breve histórico sobre BI

O termo BI já havia sido utilizado anteriormente. Como uma das primeiras referências, podemos citar a publicação de Hans Peter Luhn no *IBM® Journal of Research and Development*, em outubro de 1958, com o título “A Business Intelligence System”. Nesse artigo, Luhn discorre sobre um sistema de inteligência que estava sendo desenvolvido e que utilizava máquinas de processamento de dados para a codificação dos documentos com informações relevantes e para a definição de perfis de interesse, em pontos da ação de uma organização. Tal sistema atuaria na identificação das informações, na busca de quem precisa conhecê-las e na forma de disseminá-las de maneira eficiente.

Em meados de 1964, o desenvolvimento do IBM System 360 e de outros sistemas de mainframe tornou viável para as grandes empresas desenvolver os Sistemas de Informação de Gerenciamento (SIG), que forneciam aos gerentes

importantes relatórios periódicos estruturados, mas ainda estáticos e sem recursos de análise ou suporte interativo.

Os relatórios eram formatados com dados obtidos dos Sistemas de Processamentos Transacionais (SPT), que processavam o básico da operação de uma empresa, como estoque, pedidos, vendas etc. e ofereciam respostas às perguntas rotineiras dos gerentes, tais como: “Quanto temos deste produto no estoque?” e “Como estão as vendas em relação à meta estabelecida?”, ou seja, respostas simples processadas por rotinas sem modelos matemáticos sofisticados ou estatísticos.

Por volta de 1970, revistas e comunidades acadêmicas envolvidas começaram a publicar artigos sobre a necessidade dos sistemas de apoio à decisão, argumentando que os SIGs concentraram-se, principalmente, em gerar informações para decisões estruturadas e sugerindo que sistemas de informação para suporte a decisões mais complexas fossem criados e denominados “Sistemas de Apoio à Decisão” ou SAD.

Então, evoluções ocorreram para facilitar a vida dos gerentes médios a partir do desenvolvimento dos SADs. Um Sistema de Apoio à Decisão busca informações nos SPTs, SIGs ou em fontes externas para ajudar os gerentes de nível médio a responder questões sobre problemas que não possuem um procedimento de resolução totalmente definido, tais como: “Em determinadas circunstâncias, qual é a melhor maneira de realizar a entrega de um pedido específico?”, “Em qual tipo de pedido devemos terceirizar os serviços de entrega” e “Qual será a melhora no resultado financeiro adotando essa opção?”.

Como podemos perceber nas questões acima, um SAD deve possibilitar que os gerentes trabalhem e tomem decisões sobre grande quantidade dos dados condensados pelo sistema e, a partir dos modelos matemáticos mais avançados, analisar e tomar decisões semiestruturadas ou não estruturadas importantes para o negócio.

Apesar de um grande avanço, os gerentes sêniores e executivos apresentavam necessidades ainda mais complexas. Nesse nível de gestão, decisões não estruturadas precisam ser tomadas sobre as informações internas agregadas a uma

grande quantidade de informações externas, tais como: análises sobre ações, notícias, tendências do mercado, posicionamento dos concorrentes etc.

Para atender a tais necessidades, surgiram os SAE, Sistemas de Apoio ao Executivo, fomentados por John Rockart, da Harvard Business School, que em 1979 publicou um artigo inovador sobre esse tipo de sistema.

Em 1982, Ralph H. Sprague Jr. e Eric D. Carlson publicaram o livro *Building Efficient Effective Decision Support Systems (DDS)* e definiram os sistemas de apoio à decisão como sistemas informatizados. Neles, as decisões solucionam problemas semiestruturados ou não estruturados por interação direta com modelos de dados, realizando análises e julgamentos sobre o problema.

O conceito dos sistemas DDS foi desenvolvido com base nas necessidades e sistemas que comentamos desde a década de 1970, com o surgimento dos SIGs.

Para suprir as demandas, diversos fornecedores de software criaram ferramentas capazes de atender às necessidades de suporte à decisão.

Essas ferramentas foram possíveis graças à evolução da tecnologia. Não podemos nos esquecer de que, na década de 1960, as informações eram obtidas processando dados armazenados em arquivos. Na década de 1970, os DBMS surgiram e somente na década de 1980 os PCs e as linguagens de 4ª geração se tornaram correntes. Um dos pilares do DSS, que são as interações diretas do usuário com os modelos de dados, foram possíveis graças a tais evoluções tecnológicas.

Quando o termo BI foi criado, no final da década de 1980, havia todo sentido em definir Business Intelligence como um termo abrangente que agrupa conceitos e métodos para melhorar a tomada das decisões de negócio, usando sistemas de suporte baseados em fatos. Pois, naquele momento, existiam diversos conceitos, métodos e ferramentas no mercado, oriundos das necessidades e implementações feitas desde a década de 1970.

BI não surgiu para substituir os demais tipos dos sistemas de suporte a decisões semiestruturadas e não estruturadas (SAD, SAE e DSS), pois, como a própria definição de Business Intelligence determina, BI não se resume a um sistema.

Podemos dizer que o BI nasceu e evoluiu com base em todas as experiências anteriores e, atualmente, enquanto os demais sistemas de suporte à decisão são

aplicados em menor escala, para atender às decisões mais específicas, personalizadas de acordo com o nível gerencial ou natureza do problema, iniciativas de BI são concebidas como projetos de larga escala. O que requer grande infraestrutura para atender aos vários usuários de diversos níveis decisórios e para suportar várias tecnologias que serão utilizadas para gerar informações.

1.3 Definições atuais sobre BI

O termo Business Intelligence, tal qual o conhecemos hoje, acabou associado ao Grupo Gartner pela sua ligação com Dresner. Atualmente, podemos encontrar no site do Gartner a seguinte definição de BI: “BI é um termo guarda-chuva que inclui os aplicativos, a infraestrutura, as ferramentas e as melhores práticas que permitem o acesso e a análise de informações para melhorar e otimizar decisões e desempenho”.

A IBM define Business Intelligence (BI) em um dos seus *redbooks* como a coleta e a análise das grandes quantidades de dados para obter informações que orientem decisões de negócio estratégicas e táticas, para melhorar o desempenho de uma empresa no mercado.

Segundo a IBM, os dados podem estar relacionados a todas as facetas do negócio, como transações e distribuição demográfica dos clientes, informações financeiras da empresa, processos de fabricação, gerenciamento de inventário, tendências da indústria, transações de fornecedores e perfis dos concorrentes. Os dados de Business Intelligence são coletados de fontes internas, como sistemas de transação, processos de fabricação e registros dos clientes; e de fontes externas, como consultoria e estudos da indústria, a mídia impressa e Internet.

A SAS, importante fornecedora de soluções para Business Intelligence, definiu BI, em uma de suas publicações, como: “Uma ampla categoria de programas, aplicações e tecnologias para coletar, armazenar, analisar e fornecer acesso a dados para ajudar os usuários corporativos a tomar as melhores decisões de negócios”. Em outro trecho, completou com: “É a arte de obter uma vantagem comercial a partir de dados, respondendo a perguntas fundamentais”.

Aproveitando a definição SAS, é importante comentar que o benefício “vantagem comercial” atingido com o uso de BI, em alguns momentos, gera confusão

entre as definições de *Business Intelligence* e *Competitive Intelligence*. Apesar de ambas apoiarem a tomada das decisões, essas abordagens não são iguais, uma vez que a Inteligência Competitiva se limita a analisar e disseminar informações com foco especial nas relações com os concorrentes da empresa, enquanto o BI tem objetivos muito mais amplos.

Perceba que, na definição da IBM, o perfil dos concorrentes faz parte das informações tratadas pelo BI, ou seja, como termo guarda-chuva, podemos dizer que *Competitive Intelligence* atualmente é um subconjunto de BI.

1.4 BI vs. BA

Na mesma linha entre BI e Inteligência Competitiva, outra dúvida que se apresenta no mercado é sobre as semelhanças e diferenças entre BI e Business Analytics (BA). Utilizando o Google, podemos facilmente perceber que o interesse entre os termos está mudando, a busca por Business Analytics vem crescendo gradualmente, enquanto por Business Intelligence vem caindo bruscamente.

Não podemos analisar tal fato como um declínio do BI e muito menos da sua importância como solução para suporte a decisões, pois a busca por um termo mais antigo e consolidado tende naturalmente a diminuir, considere as pesquisas sobre EIS que incluímos na análise do Google Trends.

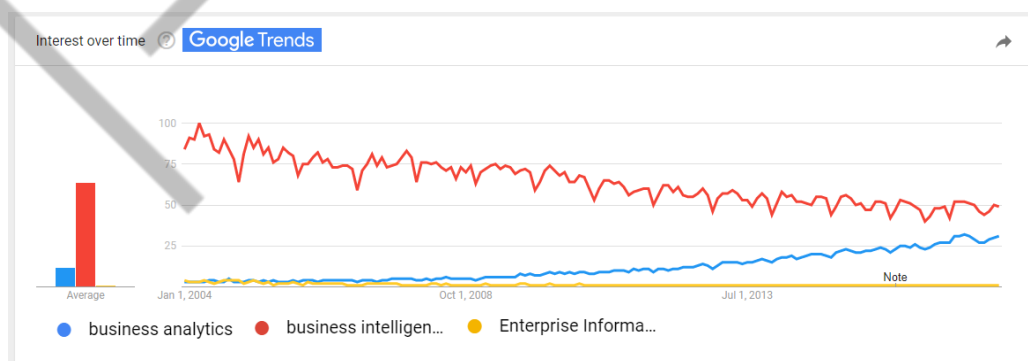


Figura 1 – Google Trends BA, BI e EI
Fonte: Adaptado por FIAP (2017)

Tal como ocorreu no surgimento das soluções BI, a implementação de Business Analytics não veio para substituir Business Intelligence e, sim, para atender às necessidades ainda ligadas a tomadas das decisões. O BI tradicional age gerando

informações sobre dados passados, dos segundos, minutos, anos ou décadas atrás, para que, por meio da base histórica, explique o que aconteceu e por que aconteceu.

Por sua vez, Business Analytics está focado em ajudar os tomadores de decisão a simular e entender possibilidades do que acontecerá.

Ambas as abordagens geram insights diferentes, mas não mais importantes. Como a partir de 2005, diversos fornecedores de soluções de BI começaram a incluir, em seus produtos, recursos e ferramentas com capacidades preditivas. Atualmente, não existe consenso no mercado se BA será um subconjunto de BI ou vice-versa.

Pois bem, se pensarmos um pouco mais sobre BA, vamos perceber que entender o que acontecerá não é uma necessidade recente. Concorde? Entender o que acontecerá, por meio de modelos preditivos estatísticos, muito menos.

Então, por que tanto interesse em Business Analytics?

A novidade está na capacidade atual de fazermos as análises, processando uma quantidade gigantesca de dados e sem profundos conhecimentos sobre estatística. Novamente, isso foi possível pelo avanço da tecnologia de hardware, redes e softwares, que, nos dias de hoje, suportam os requisitos necessários para tais esforços computacionais e, quando não *on premises*, ainda podem se valer da capacidade em *cloud*.

1.5 O conjunto que forma o BI

Como apresentamos, BI é um termo abrangente formado por um conjunto de itens que, juntos, permitem o acesso e a análise das informações para melhorar e otimizar decisões e desempenho.

Vamos entender BI pelos principais itens que o compõem:

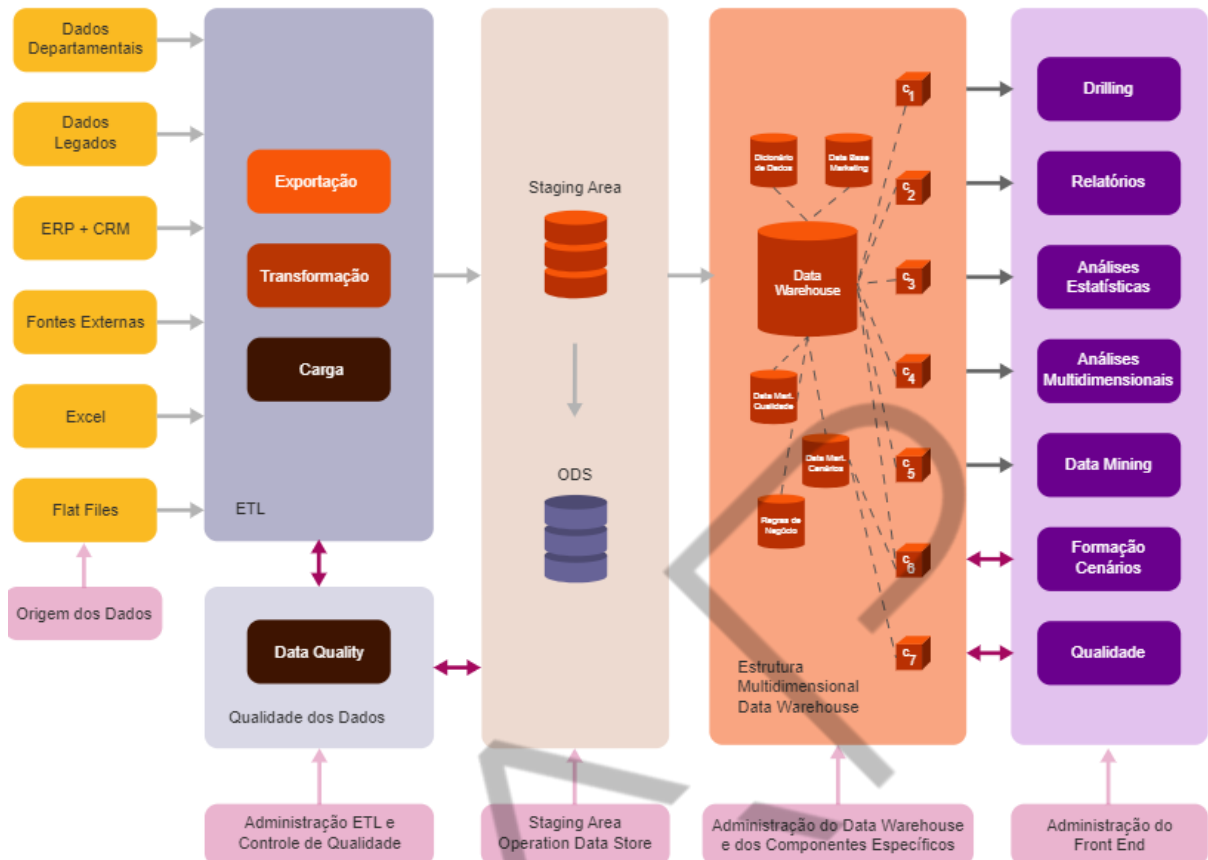


Figura 2 – Componentes do BI
Fonte: FIAP (2018)

1.5.1 Dados do BI

1.5.1.1 Fontes OLTP

Processamento das Transações On-Line (OLTP) descreve a forma como os dados são processados por um sistema informatizado. Sistemas OLTP armazenam seus dados de forma normalizada e, geralmente, processam enormes quantidades de operações CRUD realizadas pelo usuário final. Não foram criados para o BI, nesse contexto, são importantes fontes dos dados que serão utilizados pelo BI.

Podemos citar como exemplo de BI uma solução para apoio à tomada de decisões que segmenta o mercado, define os perfis dos clientes e analisa os resultados de campanhas, a rentabilidade e o risco das iniciativas do negócio. Sendo assim, podemos citar como fontes de tal solução as bases de pacotes ERP, CRM, CMS e SCM, planilhas, arquivos, sistemas departamentais e desenvolvidos sobre medida.

1.5.1.2 Fontes Externas

Dados externos são dados que não podem ser encontrados nos sistemas OLTP, mas são necessários e importantes para melhorar a qualidade das informações geradas pelo BI.

Exemplos, pesquisas sobre o mercado, informações sobre ações na bolsa, informações vindas do cadastro positivo etc. Todos seriam relevantes e complementares ao exemplo de solução de BI descrito em Fontes OLTP. Concorda?

1.5.1.3 Staging Area (SA)

A Staging Area é uma área de trabalho que recebe os dados das Fontes Internas e Externas. Os modelos de dados contidos em uma SA não precisam ser delineados seguindo uma técnica específica. Os dados são armazenados muito próximo ao seu formato original, pois não serão utilizados para consulta e análises, mas, sim, como área de cópia, limpeza e transformações feitas pelo ETL.

1.5.1.4 Operational Data Store (ODS)

Os bancos de dados operacionais são bancos de dados normalizados, desenvolvidos em algumas soluções de BI para atender às necessidades analíticas sobre processos específicos em uma empresa.

Um *Operational Data Store* propicia o armazenamento dos dados atuais, ou quase atuais, para suporte às decisões operacionais do dia a dia, por meio de consultas que retornam uma visão corporativa dos dados em nível detalhado. É orientado a assunto, integrado, porém volátil, ou seja, permite atualizações. Seus dados provêm das fontes citadas acima.

1.5.1.5 Data Warehouse (DW)

Inmon (2005), considerado o pai do Data Warehouse, define DW como um conjunto de dados de apoio às decisões gerenciais, integrado, não volátil, variável em relação ao tempo e baseado em assuntos.

- **Integrado** – Os dados são coletados a partir de uma variedade de fontes e fundidos em um todo coerente.
- **Não volátil** – Nenhum dado pode ser alterado ou excluído no DW. Qualquer consulta a um dado relativo a um período de tempo sempre produzirá o mesmo resultado; podemos complementar que nenhum dado será excluído enquanto não se tornar obsoleto para o negócio. Em soluções de BI específicas para alguns tipos de negócio, isso pode ocorrer.
- **Variável em relação ao tempo** – Todos os dados no Data Warehouse são identificados com um período de tempo particular. O DW é focado na manutenção do histórico.
- **Orientado ao assunto** – Possui dados que fornecem informações sobre um assunto específico em vez de sobre as operações em curso de uma empresa.

Segundo Kimball (2011): “Ao processo de preparar os dados de um sistema de informação operacional de forma a se ter uma fonte de informações que possam dar suporte à tomada de decisões deu-se o nome de *Data Warehousing*.”

1.5.1.6 Data Marts (DM)

São subconjuntos dos dados corporativos normalmente focados em assuntos especiais e de valor para um departamento da corporação, unidade corporativa ou conjunto de usuários. Um Data Mart é definido pelo escopo funcional que atende e não pelo seu tamanho. Geralmente, é considerado como subconjunto de um Data Warehouse.

Podemos dizer que Data Mart evoluiu a partir dos conceitos do Data Warehousing. Como a construção de um Data Warehouse pode levar muito tempo, visto que lida com informações para toda a corporação, um Data Mart pode ser considerado como um pequeno DW construído para satisfazer as necessidades de um departamento unidade ou conjunto de usuários.

Esse assunto é abordado na construção de um DW, pelas metodologias *Top-Down* (Inmon, 2005) e *Bottom-Up* (Kimball, 2011). Kimball defende que um DW seja construído a partir do desenvolvimento de DMs, ou seja, construa um Data Mart departamental e integre-o aos demais, quando estiver pronto. Inmon defende a construção do DW de uma única vez, gerando Data Marts posteriormente.

Em face do tempo de construção muito menor e o retorno de investimento mais rápido, a proposta de Kimball é a mais aceita no mercado atualmente.

2 PROCESSOS DO BI

Agora que conhecemos o papel das Fontes de Dados, da SA, do ODS, do DW e do DM em uma típica solução BI, precisamos entender como as coisas acontecem. Quais são as típicas necessidades que uma solução de BI atende? Como as informações saem das fontes e chegam ao DW e ao DM? Como essas estruturas são mantidas e administradas?

Enfim, como dados brutos se transformam em informações relevantes para a tomada de decisões?

A imagem abaixo apresenta, em alto nível, como os processos acontecem:

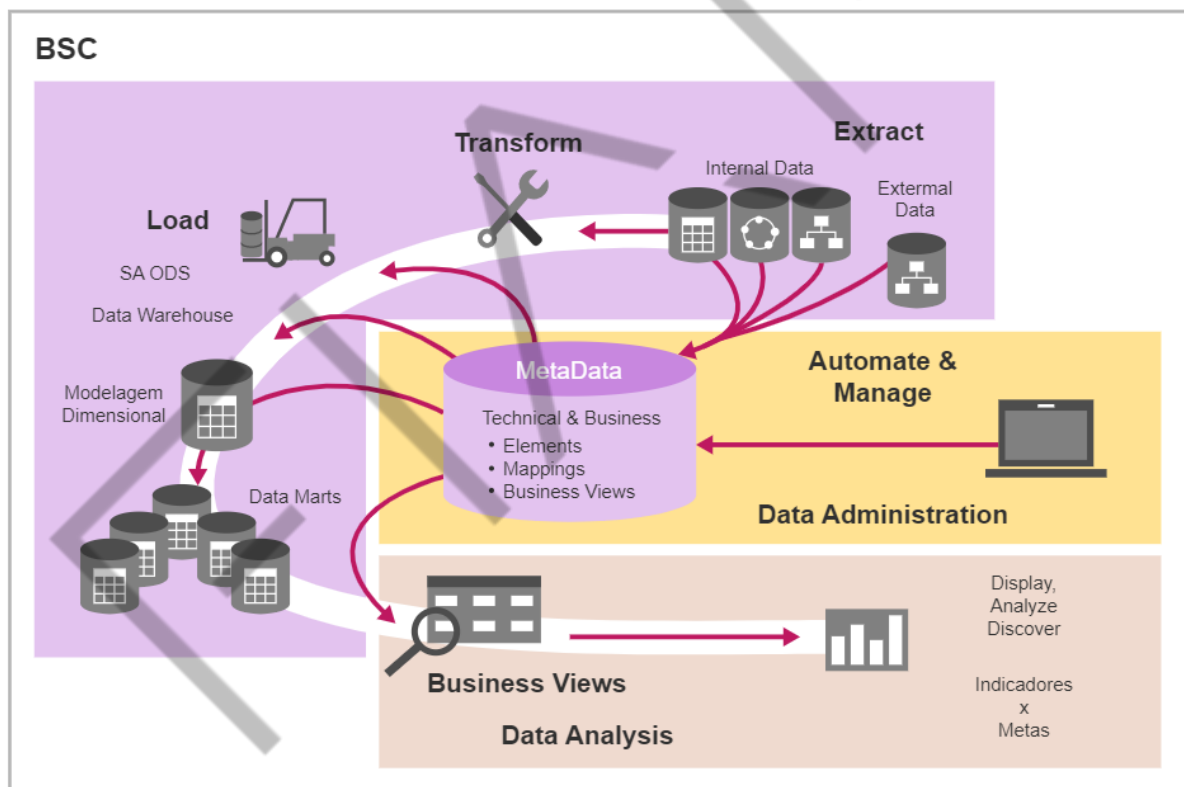


Figura 3 – Processos do BI
Fonte: Oracle, adaptado pelo autor (2018)

2.1 Business Drivers

A inteligência de negócios permeia todas as áreas de uma empresa e, como sabemos, informações mais precisas levam a decisões mais assertivas. A informação é muito significativa para o negócio e para exemplificarmos necessidades motivadoras

de BI, selecionamos duas perspectivas normalmente atendidas por soluções Business Intelligence.

- Como medir e posicionar o negócio?

O ambiente de negócios globalizado é cada vez mais competitivo. Para a sobrevivência e a entrega de bons resultados é fundamental para uma empresa aumentar suas receitas, reduzir seus custos e competir de forma cada vez mais eficaz.

- Como lidar com um ambiente de negócio tão complexo?

Atualmente, as empresas estão obtendo insumos de inúmeros fornecedores para oferecer e dar suporte às quantidades crescentes dos produtos e serviços para um conjunto cada vez maior e mais personalizado de clientes.

Para responder a tais perguntas, as soluções BI utilizam indicadores de desempenho e, frequentemente, selecionam uma ou mais abordagens de trabalho, como o Business Process Management (BPM) e o Balanced Scorecard (BSC).

O objetivo do BPM é ajudar as empresas a melhorar e otimizar suas operações em todos os aspectos do negócio, mapeando como os processos são executados, para que avanços possam ser realizados. Como os processos passam a ser conhecidos e detalhados, o BPM permite que uma solução BI aproveite os eventos que ocorrem nos processos para medir e monitorar o desempenho do negócio. Essa abordagem é mais rara de ser encontrada e geralmente é empregada por empresas que possuem alto grau de maturidade em processo.

Balanced Scorecard (BSC) é uma ferramenta de medição e gestão de desempenho que se organiza em torno de quatro perspectivas: *financeira*, *clientes*, *processos internos e aprendizagem* e *crescimento*. O BSC apresenta um caminho para a gestão do desempenho de uma empresa, capaz de ligar o plano estratégico à execução das tarefas. Essa ferramenta utiliza indicadores, definidos para cada uma das perspectivas, para o monitoramento e o controle do desempenho.

Colocando inteligência em seus negócios

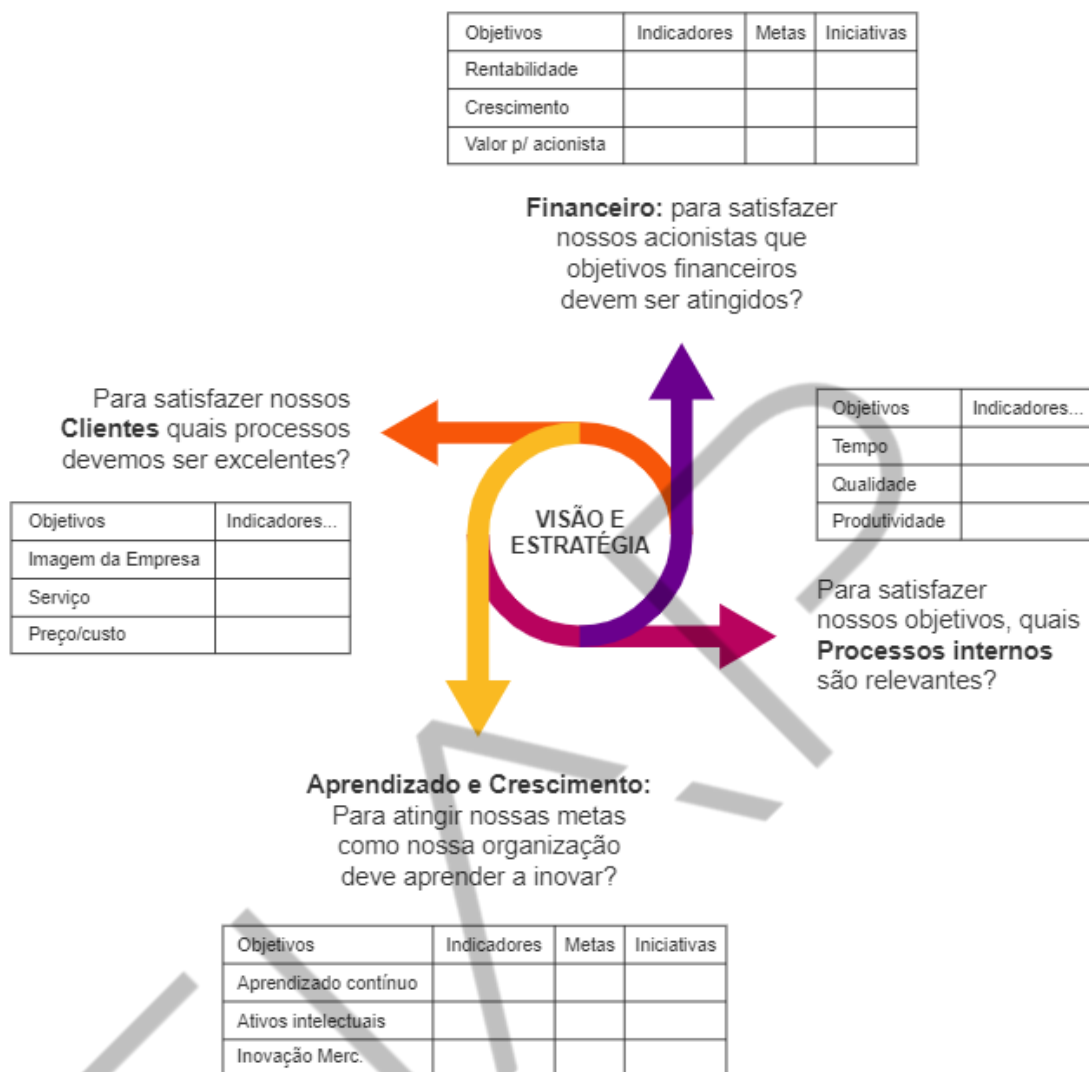


Figura 4 – Perspectivas do BSC
Fonte: Elaborado pelo autor (2018)

O Balanced Scorecard pode ser aplicado em qualquer tipo de empresa, independentemente do porte, tipo ou setor de atividade.

Os indicadores utilizados pelo BSC são do tipo quantitativo ou qualitativo, medidos com frequência e constantemente comparados às metas. Dessa forma, uma empresa que utiliza o BSC é capaz de monitorar sua performance por iniciativa que procura evoluir os resultados de uma perspectiva. Por ser mais simples, leve e orientado a estratégias, o BSC é a abordagem mais utilizada com o BI.

Para que entenda melhor como o BSC utiliza os indicadores, vamos a um exemplo sob a perspectiva do Cliente: imagine que a empresa pretende aumentar, neste ano, o tíquete médio de compras dos seus clientes (objetivo estratégico) em

10% (meta), por meio das campanhas que apresentem novos combos de produtos (iniciativa).

Agora, imagine que esse cenário é o objetivo de uma rede de supermercados, com lojas espalhadas por todo o País. Para saber se a iniciativa dará certo, será preciso calcular e monitorar o valor do tíquete médio, após a entrada em vigor da iniciativa, e comparar aos resultados obtidos em períodos semelhantes, mas anteriores ao *start* da iniciativa. Concorda?

A grosso modo, está o nosso indicador. Agora, imagine o esforço em fazer isso sem uma solução de BI, loja a loja, nota fiscal a nota fiscal, com consultas por período, por região, por estado, por cidade e por loja. Já imaginou?

Podemos dizer, a grosso modo também, que, entre outras funções, uma solução de BI automatiza o emprego do BSC.

2.2 Modelagem de dados

Existem duas abordagens para a modelagem de dados em um DW. **Inmon** defende a criação dos modelos baseados em E-R, ou modelos entidade-relacionamento. **Kimball** propõe a técnica de **modelagem dimensional**, basicamente formada por tabelas dimensionais e fato. As dimensões são tabelas que contêm os dados utilizados como perspectivas de análise, tais como a tabela data, cliente, loja, produto etc.

A tabela fato representa um processo de negócio; em nosso exemplo, a Venda é uma tabela fato. Essa tabela possui todos os itens vendidos de todas as notas fiscais de todas as lojas:

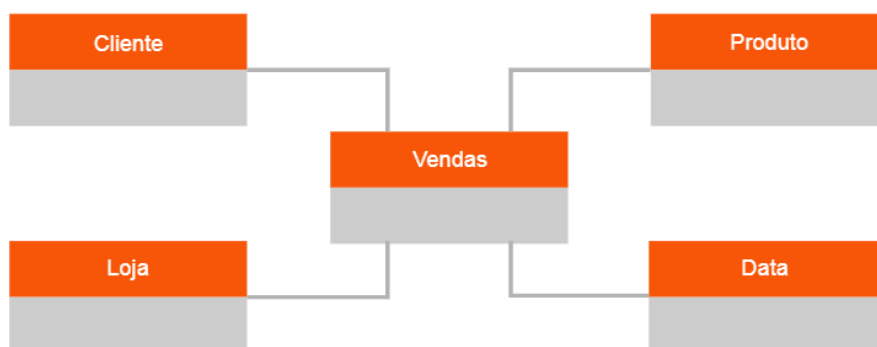


Figura 5 – Modelo dimensional de venda
Fonte: Elaborado pelo autor (2018)

Dessa forma, um usuário consegue obter informações de todas as vendas, por período (filtrando pela tabela Data), por cliente (filtrando por Cliente) e por loja (filtrando por Loja).

2.3 ETL

Extract, Transform and Load (Extração, Transformação e Carga) é o processo de extração, transformação e carga dos dados, oriundos das fontes diversas, em modelos dimensionais no DW, para que os usuários finais possam realizar consultas e tomar decisões.

- **Extração** – Obtenção dos dados brutos, das fontes internas e externas para a SA e para o ODS (quando existente).

No exemplo que utilizamos anteriormente, este passo será o responsável por buscar o valor de cada item da NF.

- **Transformação** – Limpeza dos dados extraídos, aplicação das regras de negócio e validação dos dados na SA como preparação para a carga.
- **Carga** – Inserção dos dados transformados na SA em bases de dados do DW.

A definição de dar a carga por item da NF e não pela NF como único registro diz respeito a um termo conhecido em projetos de DW, a granularidade. Podemos dizer que a granularidade é o nível de detalhe dentro de um banco de dados do DW. Quanto maior o nível de detalhe (Item da NF), menor o nível de granularidade, consequentemente, maior a quantidade de dados armazenados e maior a capacidade de consultas que podem ser atendidas.

Imagine que o usuário, após saber o valor do tíquete médio, queira saber qual produto foi mais representativo nas vendas. Se o Fato Venda tivesse informações sobre a NF, como tal interesse seria atendido? Concorda que o menor nível de granularidade sempre é o maior nível de consulta?

Voltando ao processo, as variadas fontes podem prover dados em formatos diversos, tais como dados oriundos de diferentes databases (Oracle, SQL Server, DB2 etc.) e em vários formatos físicos (planilhas, *flat files* e multimídia).

Outro ponto importante é que os Data Marts podem sofrer carga diretamente de um ETL ou de um segundo processo ETL, que buscará os dados para carga em um subconjunto de dados do DW. Isso dependerá da abordagem de construção e processos do DW.

- Manutenção

Com o passar do tempo, é natural que o DW e o DM sofram manutenções para acomodarem novos assuntos, para a remoção dos dados obsoletos e para suportar atualizações de sistemas operacionais e tecnologias envolvidas.

2.4 Administração

As soluções BI não são diferentes das demais soluções SPT, SIG, SAD, SAE e DSS nesse quesito. Todas exigem controles estruturados para manter a segurança, qualidade, integridade e relevância dos dados.

No BI, a administração das soluções pode ser otimizada com o uso de Metadados, que são dados sobre dados que armazenam informações do tipo: *de qual fonte vem o dado e onde a fonte está localizada, onde o dado está localizado na fonte, como os dados são transformados (limpeza, filtros, fórmulas, conversões, agregações etc.), para onde foi carregado, quando foi a última carga, quem acessa, quais são os padrões de acesso existentes e qualquer outra informação necessária para gerenciar a operação do DW.*

Além de auxiliar na administração, a disponibilização de Metadados aos usuários permite que confrontem a compatibilidade e a relevância dos dados que serão selecionados com os assuntos que pretendem analisar, mesmo sem grandes conhecimentos prévios sobre os dados, facilitando muito o processo de análise.

2.5 Análise dos dados

Todos os processos previamente apresentados são necessários para construir e manter as estruturas que dão suporte à análise de dados, mas uma típica solução BI possibilita aos usuários analisar os dados. Podemos dizer que as três técnicas mais comuns de análise de dados em um BI, partindo da mais simples, são:

2.6 Query and Reporting

A linguagem de consulta SQL (*Structured Query Language*) é muito utilizada para extrair de um banco de dados todas as tuplas que atendem a um conjunto específico de critérios utilizados na consulta.

Essa maneira é empregada para a construção dos relatórios e listas de registros que exibem um conjunto dos dados com características predeterminadas, geralmente para apoiar as decisões mais simples e estruturadas. O uso de SQL tem maior emprego no ODS para realizar consultas que solucionam problemas específicos do dia a dia.

No emprego no DW, as ferramentas que implementam BI aportam mecanismos geradores de SQL para que os usuários não precisem ter perfil técnico nem habilidades de análise sofisticadas.

2.7 OLAP (Processamento Analítico On-Line)

É a capacidade de manipular e analisar um grande volume dos dados através de múltiplas perspectivas e, assim, monitorar os fatos e indicadores mais relevantes da organização por meio dos painéis de controle e relatórios executivos desenvolvidos para facilitar a visualização, o entendimento dos fatos e a tomada de decisões. Utiliza técnicas específicas, como ranking, comparações, *pivot tables*, *drill down* e *roll up*, *slice* e *dice*, entre outras.

Esta é uma das formas de análise feitas sobre os dados transformados pelo ETL e carregados em múltiplas dimensões, por assunto e pontos de vista do negócio. Essa técnica é usada, por exemplo, para obter informações em um BI sobre como os eventos de vendas de uma empresa ocorrem ao longo de intervalos de tempo.

Ferramentas OLAP permitem que os próprios usuários explorem, analisem e respondam a perguntas do tipo:

- Qual filial vendeu mais em valores absolutos este ano?
- Qual filial vendeu mais em valores absolutos no primeiro trimestre deste ano?

- Qual a participação do Departamento de Eletrodomésticos em valores absolutos no primeiro trimestre deste ano, na filial que mais vendeu neste período?

Utilizando as perguntas na ordem que foram feitas, da primeira para a segunda, o usuário fez uma operação conhecida como *drill down*, ou seja, quando recebeu a resposta, optou por navegar de uma resposta menos detalhada em busca de uma mais detalhada.

Ao obter a segunda resposta, o usuário empregou uma segunda técnica, conhecida como *slice*, ou seja, de todos os departamentos, o usuário selecionou apenas eletrodomésticos, fatiando o conjunto de dados resultante para analisar apenas a fatia que lhe é interessante, neste caso, o Departamento de Eletrodomésticos.

Essa facilidade de navegação provida pelo OLAP para que os próprios usuários escolham os caminhos de análise livremente é possível porque o OLAP usa tabelas do DW e do DM para permitir consultas multidimensionais das grandes quantidades de dados. Em nosso exemplo, o assunto é vendas, e as dimensões utilizadas foram Filial, Data e Departamento.

Os cubos em OLAP podem ser implementados de quatro formas diferentes:

- **ROLAP** (*Relacional On-Line Analytical Processing*): os dados são armazenados em banco de dados relacionais e cada ação de *Slice/Dice* é equivalente a adicionar uma condição na cláusula WHERE em uma *query* SQL. Implementação contida nas figuras deste documento.
- **MOLAP** (*Multidimensional On-Line Analytical Processing*): os dados são armazenados em cubos dimensionais, em formatos proprietários das ferramentas, e não em banco de dados relacionais. O usuário monta os cubos e manipula os dados diretamente no servidor OLAP.
- **HOLAP** (*Hybrid On-Line Analytical Processing*): combina as vantagens de MOLAP e ROLAP. Para informações do tipo síntese, HOLAP utiliza cubos dimensionais para um retorno do resultado mais rápido. Quando a análise requer mais informação, o HOLAP pode complementar o cubo multidimensional buscando dados no banco de dados relacional.

- **DOLAP** (*Desktop On-Line Analytical Processing*): ferramentas que disparam uma consulta de uma estação cliente para o servidor, que, por sua vez, retorna enviando um cubo de volta, para que possa ser analisado pelo usuário.

2.8 Data Mining

Esta é uma técnica de análise para fazer descobertas automáticas sobre padrões ocultos e relacionamentos desconhecidos de similaridade ou discordância entre dados, dentro de um grande volume de dados.

O *Data Mining* utiliza algoritmos complexos e técnicas de Inteligência Artificial para fazer tais descobertas e transformá-las em informações úteis para a tomada de decisão e/ou a avaliação de resultados. Atualmente, um emprego muito comum desta técnica é na prevenção de fraude, segmentação de clientes e previsão de comportamento.

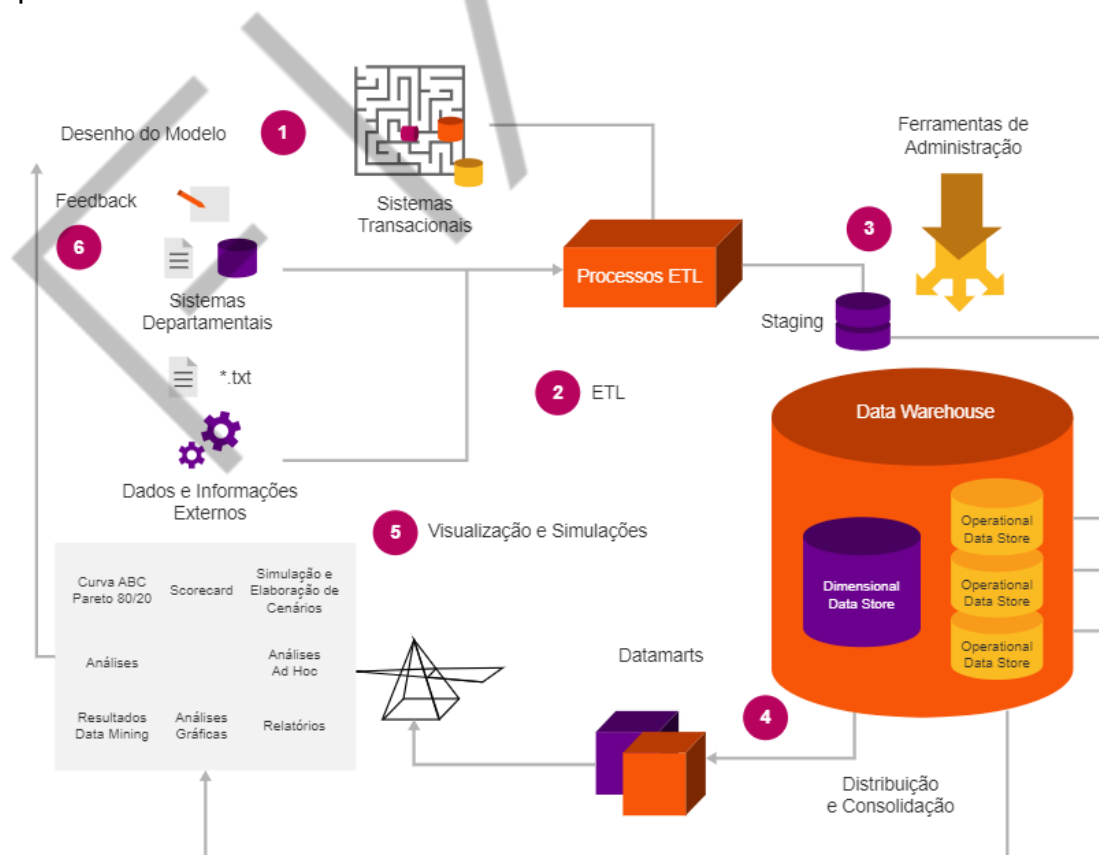


Figura 6 – Visão geral do BI
Fonte: FIAP (2018)

CONCLUSÃO

O objetivo deste conteúdo foi falar o básico e essencial sobre Business Intelligence. Apresentamos um breve histórico sobre BI, as definições atuais e discutiremos sobre algumas diferenças entre termos parecidos do mercado.

Foram elencados também os principais componentes e processos que compõem uma solução de BI. Mostramos os principais motivadores e uma abordagem baseada em BSC para exemplificar como uma solução de Business Intelligence propicia a uma grande corporação, como uma rede de supermercados, monitorar indicadores e compará-los a metas para saber se as iniciativas estão atingindo os objetivos estratégicos.

Lembro que as várias definições do termo Business Intelligence utilizam as palavras guarda-chuva, abrangente e conjunto. Sendo assim, existem diversos assuntos sobre os aplicativos, a infraestrutura, as ferramentas, os conceitos, os métodos e as melhores práticas que formam uma solução de BI.

Após esta imersão no mundo do Business Intelligence, vamos avançar para um conceito abordado neste capítulo: Modelagem Dimensional e suas técnicas!

REFERÊNCIAS

BALLARD, Chuck et al. **Dimensional Modeling**: in a Business Intelligence environment. USA: IBM-International Technical Support Organization, 2006.

BARBIERI, C. **BI – Business Intelligence**: modelagem e tecnologia. Rio de Janeiro: Axcell Books, 2001.

BARBIERI, C. **BI2 – Business Intelligence**: modelagem e qualidade. São Paulo: Campus, 2011.

GONÇALVES, M. **Extração de dados para Data Warehouse**. Rio de Janeiro: Axcell Books, 2003.

HAHN, Seungrahn et al. **Capacity planning for Business Intelligence applications**: approaches and methodologies. USA: IBM Redbooks, 2000.

INMON, W. H. **Building the Data Warehouse**. New Jersey: John Wiley & Sons, 2005.

KAPLAN, Robert S.; NORTON, David P. **Mapas estratégicos**: convertendo ativos intangíveis em resultados tangíveis. 7. reimpressão. Tradução de Alfonso Celso de Cunha Serra. Rio de Janeiro: Elsevier, 2004.

KIMBALL, R.; CASERTA, J. **The Data Warehouse ETL Toolkit**: practical techniques for extracting, cleaning, conforming and delivering data. Nova Jersey: John Wiley & Sons, 2011.

KIMBALL, R.; ROSS, M. **The Data Warehouse Toolkit**: the complete guide to Dimensional Modeling. Nova Jersey: John Wiley & Sons, 2011.

LAUDON, Kenneth C.; LAUDON, Jane P. **Sistemas de Informações Gerenciais**. 7. ed. São Paulo: Pearson, 2007.

MACHADO, F. N. R. **Tecnologia e Projeto de Data Warehouse**: uma visão multidimensional. São Paulo: Érica, 2006.

MARTINS, Tomas Sparano et al. **Incrementando a estratégia**: uma abordagem do Balanced Scorecard. Curitiba: Intersaberes, 2010.

NELSON, G. **Introduction to the SAS® 9 Business Intelligence Platform**: a tutorial. In: SAS GLOBAL FORUM. 2007. Disponível em: <https://www.researchgate.net/publication/237750763_Introduction_to_the_SASR_9_Business_Intelligence_Platform_A_Tutorial>. Acesso em: 14 abr. 2020.

POWER, D. J. **A brief history of decision support systems**. 2007. Disponível em: <<http://dssresources.com/history/dsshhistory.html>>. Acesso em: 13 out. 2017.

REINSCHMIDT, J.; FRANCOISE, A. **Business Intelligence Certification Guide**. USA: IBM International Technical Support Organization, 2000.

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

TURBAN, Efraim et al. **Business Intelligence:** a managerial approach. New Jersey: Upper Saddle River & Pearson Prentice Hall, 2008.

TURBAN, Efraim et al. **Business Intelligence:** um enfoque gerencial para a inteligência do negócio. Porto Alegre: Bookman, 2009.

WATSON, H. J.; WIXOM, B. H. The current state of Business Intelligence. **Computer**, v. 40, n. 9, p. 96-99, 2007.

EXEMPLO

GLOSSÁRIO

Gartner Group	Gartner é uma empresa de consultoria fundada em 1979 por Gideon Gartner.
SPT	Os Sistemas de Processamentos Transacionais fazem o processo básico da operação no dia a dia de uma empresa, como estoque, pedidos, vendas etc.
SIG	Os Sistemas de Informação de Gerenciamento (SIG) fornecem aos gerentes importantes relatórios periódicos estruturados, mas ainda estáticos e sem recursos de análise ou suporte interativo.
IBM System 360	O IBM System/360 (S/360) é uma família de <i>mainframes</i> lançada pela IBM em 7 de abril de 1964. Foi a primeira família de computadores a fazer uma distinção clara entre a arquitetura e a implementação, permitindo que a IBM lançasse um conjunto de projetos compatíveis em várias faixas de preço.
SAD	Os Sistemas de Apoio à Decisão ou SAD buscam informações nos SPTs, SIGs ou em fontes externas para ajudar os gerentes de nível médio a responder questões sobre problemas que não possuem um procedimento de resolução totalmente definido.
SAE	Os Sistemas de Apoio Executivo são sistemas do nível estratégico desenvolvidos para apoiar executivos nas decisões não estruturadas que precisam ser tomadas sobre informações internas agregadas a uma grande quantidade de informações externas, tais como: análises sobre ações, notícias, tendências do mercado, posicionamento dos concorrentes etc.
DDS	Decision Support Systems são sistemas informatizados, nos quais os decisores solucionam problemas

	semiestruturados ou não estruturados, por interação direta com os modelos de dados, realizando análises e julgamentos sobre o problema.
DBMS	Database Management System (DBMS) ou Sistema Gerenciador de Banco de Dados é um software que interage com usuários finais, outras aplicações e o próprio banco de dados para capturar e analisar dados.
IBM	International Business Machines (IBM) é uma empresa dos Estados Unidos voltada à área de Informática. A empresa é uma das poucas do setor de Tecnologia da Informação (TI) com uma história contínua que remonta ao século XIX.
SAS	SAS Institute é um desenvolvedor multinacional americano de software de análise baseado em Cary, Carolina do Norte. O SAS desenvolve e comercializa um conjunto de softwares de análise (também chamado de SAS – Sistema de Análise Estatística) que ajudam a acessar, gerenciar, analisar e divulgar dados para auxiliar na tomada de decisões.
Competitive Intelligence	A Inteligência Competitiva é o processo contínuo de monitoramento e análise estratégica dos cenários e conjunturas mercadológicas em que determinada empresa está inserida. A Inteligência Competitiva (IC) busca identificar tendências do mercado, desenvolver análises estratégicas, descobrir oportunidades e mapear riscos por meio de metodologias científicas.
<i>Business Analytics</i>	A Análise de Negócios (BA) refere-se às habilidades, tecnologias e práticas para a exploração iterativa contínua e a investigação do desempenho das empresas passadas para obter informações e impulsionar o planejamento de negócios.

Google Trends	Google Trends é uma ferramenta do Google que mostra os mais populares termos buscados em um passado recente.
On premises	Software local instalado e executado em computadores nas instalações da organização que usa o software.
Cloud computing	O conceito de computação em nuvem refere-se à utilização da memória e da capacidade de armazenamento e cálculo de computadores e servidores compartilhados e interligados por meio da Internet.
OLTP	Processamento de Transações On-Line (OLTP) descreve a forma como os dados são processados por um sistema informatizado. Sistemas OLTP armazenam seus dados de forma normalizada e, geralmente, processam enormes quantidades de operações CRUD, realizadas pelo usuário final.
CRUD	CRUD (acrônimo de <i>Create, Read, Update e Delete</i>) equivale às quatro operações básicas utilizadas em bases de dados relacionais (RDBMS) ou em interface para utilizadores, para criação, consulta, atualização e destruição de dados.
ERP	O <i>Enterprise Resource Planning</i> é uma plataforma de software desenvolvida para integrar os diversos departamentos de uma empresa, possibilitando a automação e o armazenamento de todas as informações do negócio.
CRM	<i>Customer Relationship Management</i> é um termo em inglês criado para definir toda uma classe de sistemas de informações ou ferramentas que automatizam as funções de contato com o cliente.
CMS	O Sistema de Gerenciamento de Conteúdo é um aplicativo usado para criar, editar, gerenciar e publicar conteúdo de

	forma consistentemente organizada, permitindo que o mesmo seja modificado, removido e adicionado com facilidade. CMSs são com frequência utilizados para armazenar, controlar e prover documentação empresarial, tais como notícias, artigos, manuais de operação, manuais técnicos, guias etc. O conteúdo pode incluir imagens, áudios, vídeos, documentos e conteúdo Web.
SCM	Gestão da Cadeia de Suprimentos de Negócios ou <i>Supply Chain Management</i> é o gerenciamento de uma rede interligada de negócios envolvidos na provisão final de produtos e serviços requeridos por clientes finais. Como tal gerenciamento é complexo, produtos baseados em software foram desenvolvidos.
SA	A Staging Area é uma área de trabalho que recebe os dados das Fontes Internas e Externas para serem transformados pelo ETL, antes de serem carregados para o DW.
ODS	Os bancos de dados operacionais são bancos de dados normalizados desenvolvidos em algumas soluções de BI para atender às necessidades analíticas sobre processos específicos em uma empresa.
DW	Data Warehouse é um conjunto de dados de apoio às decisões gerenciais, integrado, não volátil, variável em relação ao tempo e baseado em assuntos.
DM	Data Mart são subconjuntos de dados corporativos geralmente focados em assuntos especiais e de valor para um departamento da corporação, unidade corporativa ou conjunto de usuários. Um Data Mart é definido pelo escopo funcional que atende e não pelo seu tamanho. Normalmente, é considerado como subconjunto de um Data Warehouse.

BPM	Business Process Management é um conceito que une Gestão de Negócios e Tecnologia da Informação com foco na otimização dos resultados das organizações por meio da melhoria dos processos de negócio.
BSC	Balanced Scorecard (BSC) é uma ferramenta de medição e gestão de desempenho que se organiza em torno de quatro perspectivas: financeira, clientes, processos internos e aprendizagem e crescimento.
OLAP	É a capacidade de manipular e analisar um grande volume de dados através de múltiplas perspectivas e, assim, monitorar os fatos e indicadores mais relevantes da organização por meio de painéis de controle e relatórios executivos desenvolvidos para facilitar a visualização, o entendimento dos fatos e a tomada de decisões.
ETL	Extract, Transform and Load (Extração, Transformação e Carga) é o processo de extração, transformação e carga dos dados oriundos de fontes diversas em modelos dimensionais no DW, para que os usuários finais possam realizar consultas e tomar decisões.
SQL	<i>Structured Query Language</i> é um idioma-padrão para armazenar, manipular e recuperar dados em bancos de dados.
<i>Pivot tables</i>	Uma tabela dinâmica é uma tabela que resume dados em outra tabela feita aplicando uma operação, como classificar, calcular a média ou somar dados na primeira tabela, tipicamente incluindo o agrupamento dos dados.
<i>Drill down</i>	Ato de fazer uma exploração em diferentes níveis de detalhe ou hierarquias de informação de uma dimensão, partindo da menos detalhada para a mais detalhada.
<i>Roll up</i>	É o ato inverso ao <i>Drill down</i> , ou seja, uma exploração em diferentes níveis de detalhe ou hierarquias de informação

	de uma dimensão, partindo da mais detalhada para a menos detalhada.
<i>Slice</i>	<i>Slice</i> é um filtro que permite ver os dados de diferentes pontos de vista e apresenta dados de uma única dimensão de um cubo.
<i>Dice</i>	<i>Dice</i> é um filtro que permite ver os dados de diferentes pontos de vista e mostra um subcubo ou intersecção de vários <i>slices</i> .