

DISCIPLINA DE BUSINESS INTELLIGENCE



Sumário

UNIDADE I - O PODER DA INFORMAÇÃO	4
Introdução.....	4
O USO DA INFORMAÇÃO NO MUNDO ORGANIZACIONAL	5
1. O que são Indicadores.....	6
2. Indicadores de desempenho	8
3. Estudo de caso	14
4. Considerações finais.....	15
Referências	18
UNIDADE II - BUSINESS INTELLIGENCE	19
Introdução.....	19
1. Conceito de Business Intelligence (BI)	20
2. Data Warehouse (DW)	27
3. Projeto de Business Intelligence (BI)	29
4. Configurações diversas	33
5. Inicialização das ferramentas	35
6. Big Data	41
7. Considerações finais.....	43
Referências	46
UNIDADE III - MODELO DE PREPARAÇÃO DE UM DATA WAREHOUSE	47
Introdução.....	47
1. Cultura e vantagem de um Data Warehouse.....	48
2. Data Warehouse.....	49
3. Considerações finais.....	66
Referências	69
UNIDADE IV – O PODER DA FERRAMENTA ETL	70
Introdução.....	70
1. A ETL.....	71
2. Papel e responsabilidade da ETL.....	72
3. Problemas de ETL.....	74
4. Categorias de ferramentas de ETL.....	75
5. Query cache	83
6. Considerações finais.....	85
Referências	88

UNIDADE I - O PODER DA INFORMAÇÃO

Introdução

Na busca pela excelência empresarial, a informação se tornou uma ferramenta indispensável para apoiar as empresas a atingirem seus resultados e seus objetivos econômicos e sociais. Além de sistemas de informação eficientes, na busca por resultados satisfatórios, os caminhos da gestão necessitam de agilidade na tomada de decisão.

Buscando a excelência empresarial, certas necessidades são impostas para gerir melhor a gama de informações, para que tudo rode conforme a sua necessidade, visando sempre os prazos dos indicadores financeiros, lembrando que estes dados são para uma análise sistêmica das estratégias da organização, as quais são alinhadas aos indicadores.

Dentro do conceito de sistema, a medição de desempenho evidencia a necessidade de estabelecer critérios de mensuração que devem compor o quadro de indicadores financeiros e não financeiros que possam traduzir a estratégia da organização.

No processo gerencial, a escolha dos indicadores tem a necessidade de equalizar os processos, e é possível utilizar os esquemas de gestão para auxiliar na avaliação dos resultados.

Os indicadores devem representar tendências e comparações de forma a abranger e enfatizar de forma precisa os fatores de quantidade, qualidade, valor financeiro e tempo.

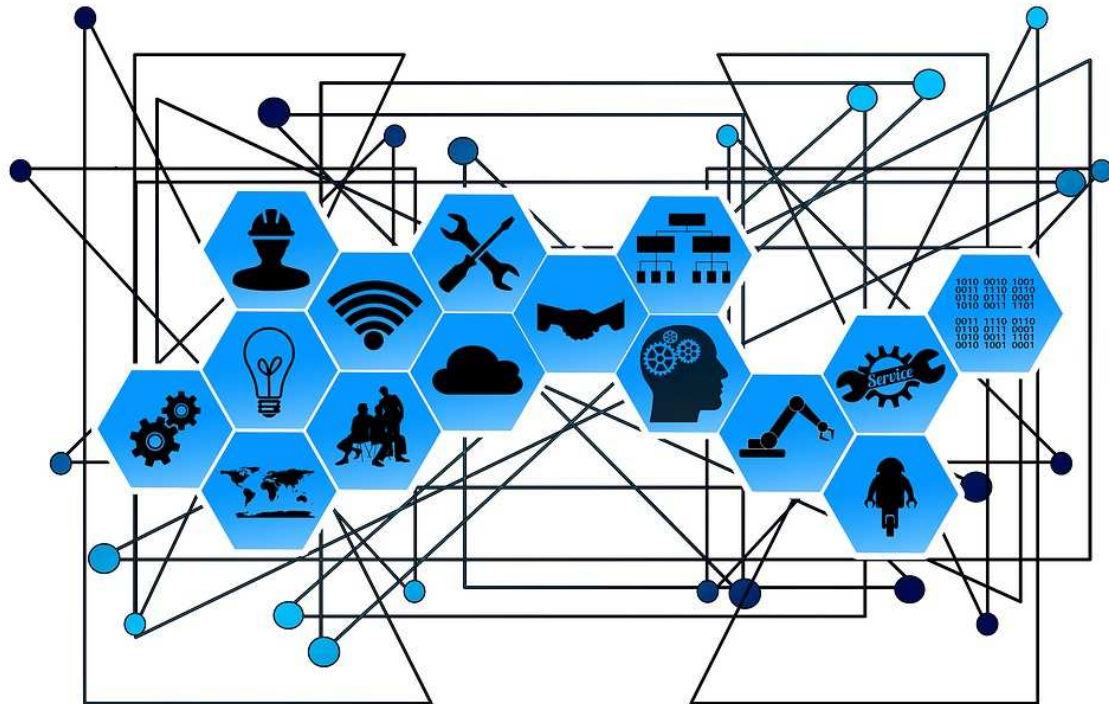
Os critérios de desempenho para revelar a grandeza e monitorar os objetivos dos indicadores, podem ter suas medidas determinadas por meio de indicadores de eficácia, eficiência, qualidade, produtividade, inovação, custo e lucratividade.

Os critérios de mensuração associados com variáveis importantes para o desempenho empresarial fortalecem a eficiência empresarial. O mesmo ocorre com as variáveis de eficácia, qualidade, produtividade, lucratividade e as demais. Com esse alinhamento, é possível melhorar o desempenho e contribuir favoravelmente no resultado total da organização.

A partir desse contexto o objetivo desta unidade é mostrar o papel dos indicadores de desempenho no processo de gestão, bem como, evidenciar algumas de suas classificações.

Ótimo estudo!

O USO DA INFORMAÇÃO NO MUNDO ORGANIZACIONAL



Disponível em: <https://pixabay.com/pt/rede-pontos-linhas-interfaces-2496193/>

A busca pela eficiência nos resultados empresariais apresenta a necessidade de melhora na qualidade e pontualidade das informações gerenciais. Isso permite ultrapassar a visão de curto prazo apenas de indicadores financeiros para uma análise do sistema de estratégias organizacionais, alinhadas aos indicadores de performance financeiros e outros.

Os indicadores têm o objetivo de medir e avaliar o andamento da empresa, lembrando da situação financeira, e a eficiência das estratégias adotadas para melhor satisfação do cliente. Para a adoção de processos e estratégias utilizamos os indicadores de desempenho para preparar as informações fornecidas ao gestor, para que o mesmo tenha facilidade na análise e nas tomadas de decisões.

Conhecidos como Key Performance Indicator (KPIs), este indicador é utilizado entre as empresas e suas concorrentes, para que seja possível extrair uma análise de desempenho e desenvolvimento dos setores, permitindo assim obter indicadores sobre os setores das empresas.

De acordo com as informações disponíveis no site Novo Negócio ([2016], on-line):

os indicadores de desempenho são utilizados para avaliar e medir o desempenho da empresa, desde a sua saúde financeira, eficiências

das estratégias adotadas até a satisfação do cliente. Para os processos e estratégias, os indicadores de desempenho são utilizados para preparar informações para que o gestor tenha facilidade na análise e nas decisões a serem tomadas. Também conhecidos como KPIs (*Key Performance Indicator*), os indicadores podem ser utilizados entre uma empresa e suas concorrentes, sendo possível obter uma análise sobre o desenvolvimento e desempenho dos setores, além de permitir a avaliação dos setores da empresa com melhor e pior desempenho.

1. O que são Indicadores



Disponível em: <https://pixabay.com/photo-1925495/>

Iniciando com a famosa citação de Peter Ferdinand Drucker que diz que “O que não é medido não pode ser gerenciado”, o renomado escritor, professor e consultor austríaco considerado até hoje o pai da administração moderna apresenta a essência dos indicadores e a forma de efetuar a leitura sobre eles.

A sigla KPI, do inglês “*Key Performance Indicator*” é traduzido literalmente como: “Indicador Chave de Desempenho”.

A finalidade dos KPIs é a de tornar possível utilizar ferramentas de gestão e administração largamente por empresas do mundo todo para medir e avaliar o desempenho de seus processos e permitir gerenciá-los da maneira mais eficaz e eficiente possível. O objetivo visa a conquista de metas e objetivos previamente estabelecidas pelas organizações com os respectivos resultados alcançados.

Saiba mais ou atenção



Disponível em: <https://pixabay.com/photo-1853619/>

Cada objetivo estratégico é construído com um conjunto de medidas de desempenho (Indicadores Chave de Desempenho ou *Key Performance Indicators* - KPI) que permitem monitorar o atingimento destes objetivos. Esses objetivos devem estar em alinhamento com a Administração da empresa e ter o conhecimento adequado de cada área que se deseja atingir objetivos?

Os tipos de indicadores de desempenho são diversos, cada um com uma finalidade e circunstâncias diferentes. Podem ser quantitativos ou qualitativos, dependendo da intenção do administrador e dos tipos de KPI escolhidos. Podem tanto avaliar os processos em números como determinar a qualidade que estão sendo executados.

Além de poderosas ferramentas da gestão de processos, os KPIs também funcionam como meio de comunicação organizacional. É por meio deles que o desenvolvimento das empresas é compartilhado com os parceiros e funcionários dos diversos níveis hierárquicos e organizacionais.

Dessa maneira, tipos diferentes de indicadores ajudam na transmissão síncrona da missão, visão e valores da empresa a todos os funcionários, com o intuito de mantê-los alinhados, integrados e fazendo com que eles entendam a importância de suas funções dentro do coletivo e corporativo.

Fonte: o autor.

Por dentro dos fatos



Disponível em: <https://pixabay.com/photo-2328525/>

A Oracle possui um modelo da experiência do cliente juntamente com o KPI. Ela define a experiência do cliente da Oracle, nominado como CX (CustomereXperience). O Valor da Equação é o método estratégico para identificar e medir o valor da experiência do cliente organizacional. Esse modelo foi originalmente documentado em "CX Métricas e KPIs - Setembro de 2012". Usando Valor da Equação da experiência do cliente, é possível identificar os KPIs específicos mais adequados para cada negócio; apresenta como é possível maximizar o valor do investimento da experiência do cliente e o trabalho para desenhar um roteiro de implementação que irá garantir o sucesso.

Fonte: adaptado de Oracle (2016, on-line).

2. Indicadores de desempenho



Disponível em: <https://pixabay.com/photo-816492/>

Os indicadores de desempenho são métricas que auxiliam na avaliação do desempenho de situações que necessitam ser avaliadas, seja setor, atividade, setores ou até mesmo comparar desempenhos para encontrar possíveis problemas e soluções. Segundo Macedo-Soares e Ratton (1999) é dada, mundialmente, grande importância à questão da medição de desempenho

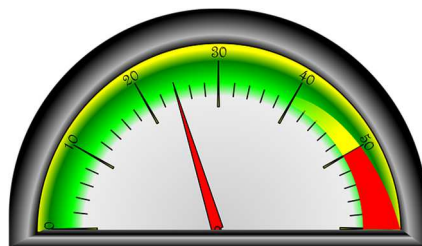
Hoje, além dos dados (números) fornecidos pelos indicadores de desempenho é possível encontrar sistemas em forma de gráficos e muito intuitivos, o que permite uma rápida e fácil visão e decisão com mais exatidão, seguindo o apontamento do indicado. É por meio das métricas que o gestor poderá ter assertividade em suas decisões, de forma muito mais rápidas, mensurando rapidamente e com solidez.

2.1 Tipos de Indicadores de Desempenho: Processo e Estratégia

A grande charada para o uso desta ferramenta é saber diferenciar os dois tipos de indicadores de desempenho, o processo e a estratégia. De acordo com as informações disponíveis no site Novo Negócio ([2016], on-line):

É importante diferenciar os dois tipos de indicadores de desempenho: processo e estratégia. Os indicadores de estratégias têm por função analisar e verificar se os objetivos das metas estão sendo cumpridos, pela direção que estão caminhando. Os indicadores de estratégia têm o objetivo de analisar e verificar se as metas traçadas estão sendo atingidas pela direção. Para facilitar no processo de tomada de decisão, conforme (KAPLAN; NORTON, 2007) o mais utilizado é o método Balanced Scorecard (Indicadores Balanceados de Desempenho), pois pela sua capacidade de integrar as ações Estratégicas às Operacionais e Organizacionais de uma empresa, possibilita a sincronia na definição das estratégias das empresas, com a gestão do negócio, com o gerenciamento dos serviços e tendo o foco total na qualidade. O foco dos indicadores de processos é a avaliação e análise de como a tarefa está sendo executada, mensurando o desempenho e resultados. Este indicador é medido através de dados e índices, normalmente números, facilitando a tomada de decisões e ajustes.

2.2 Como usar os indicadores de desempenho



Disponível em: <https://pixabay.com/photo-161160/>

Para entender como utilizar os indicadores de desempenho, existem diversas opções de indicadores e para utilizá-los corretamente, maximizando o resultado e a eficiência são apresentados três passos, seguem:

- Tenha sempre o objetivo bem definido:
A falta de objetivo específico impede a utilização dos indicadores de desempenho corretos. São dezenas de indicadores que podem ser utilizados, a questão importante a ser levada em consideração é qual ou quais devem ser levados em consideração. Para isso é necessário ser específico e saber exatamente o que se deseja descobrir.
- Se a busca for por alguma falha no sistema de faturamento, os indicadores de lucratividade e rentabilidade podem ser as melhores opções.
- Se for referente aos clientes, os indicadores de qualidade e valor são boas opções.

A contradição dos indicadores de desempenho

Indicadores de desempenho se contradizem: para iniciar a análise é importante escolher entre os indicadores de desempenho. O cenário ideal é analisar dois ou três indicadores sendo eles para cada tipo de análise específica. Onde se o foco se tornar a análise de vários tipos de indicadores pode se fazer com que a análise seja comprometida ao ponto de impedir a tomada de decisão.

Análise dupla

“A análise dupla é importante, pois dão mais segurança aos indicadores analisados. A decisão dos indicadores que serão analisados, necessitam de maior segurança. Para isto, faça uma análise dupla” (NOVO NEGÓCIO, [2016], on-line).

Com dois indicadores de desempenho em mãos para medir a qualidade da prestação de serviços, é viável ter uma decisão por alguma mudança, entretanto, com outros dois indicadores parecidos, para avaliar mais especificamente aquele processo, faça novamente a análise e verifique a assertividade. Caso exista discordância entre as análises e as necessidades de se tomar uma decisão, é o momento de ter muito cuidado em sua decisão e ter os olhos voltados aos detalhes a se efetuar as correções de se tomar cuidado e se ter muita calma para fazer as correções (NOVO NEGÓCIO, [2016], on-line).

2.3 Tipos de Indicadores de Desempenho de Processos



Disponível em: <https://pixabay.com/pt/gr%C3%A1fico-crescimento-finan%C3%A7as-lucros-163509/>

Segue uma relação com dez tipos de indicadores de desempenho mais utilizados nas análises da gestão empresarial:

- **Indicadores de Eficácia:** dizem respeito aos resultados alcançados em relação aos resultados pretendidos. É possível dizer que o maior foco dos indicadores de eficácia está no produto, o resultado obtido e os lucros que vêm deles. Ou seja, mede a melhor forma de tomar a decisão, agindo e visando os impactos para a empresa.
- **Indicadores de Eficiência:** levam em consideração os resultados obtidos em relação aos recursos empregados. Têm como objetivo a redução de custos, mensura os recursos aplicados. No sentido mais

puro, a eficiência consiste em tomar a decisão de fazer da melhor maneira possível, sem utilizar muitos recursos.

- **Indicadores de Capacidade:** trazem a relação de quantidade que se pode produzir pelo tempo de produção. Exemplo: uma fábrica de roupas produz 100 mil unidades de calças jeans por mês.
- **Indicadores de Rentabilidade:** são o lucro em relação ao investimento realizado pela empresa, ou seja, é a análise da porcentagem entre o investimento feito e o lucro. Nesse cenário, uma empresa que investe cerca de 500 mil reais e tem 100 mil de lucro, apresenta uma rentabilidade de 20%.
- **Indicadores de Competitividade:** são a relação de dados entre a empresa e a concorrência, assim como o Market Share. Como o nome sugere, este indicador tem o objetivo principal de mensurar a relação da empresa com as concorrentes diretas.
- **Indicadores de Lucratividade:** são o percentual de lucro em relação às vendas. Não é muito difícil entender que esses indicadores medem a relação percentual entre as vendas e o lucro. Se uma empresa vender 550 mil reais em produtos e o lucro for de 55 mil, a lucratividade girará em torno dos 10%.
- **Indicadores de Qualidade:** levam em consideração tudo o que foi produzido em relação às saídas adequadas ou em conformidade. Exemplo: uma empresa que produz calças jeans produz cerca de 950 mil peças de primeira linha por mês ao uso de uma produção total de 1 milhão, com isto a porcentagem de qualidade é de 95%.
- **Indicadores de Efetividade:** são a junção ou a verificação dos resultados da conjugação entre os indicadores de eficácia e eficiência.
- **Indicadores de Valor:** são o valor que indica a relação entre o valor percebido ao receber um produto e o valor real despendido para o seu recebimento.
- **Indicadores de Produtividade:** são a relação do esforço de trabalho com os recursos gastos para o resultado. Utilizado para avaliação de desempenho de funcionários. Aqui, o administrador deve fazer a leitura da relação entre os recursos utilizados para determinado trabalho e as saídas geradas por ele. Exemplo: uma costureira consegue montar quatro calças jeans em uma hora de trabalho, enquanto outro monta somente duas, de igual dificuldade e no mesmo período. Isso significa que a primeira é mais produtiva.

Atualmente, os diversos tipos de indicadores de desempenho à disposição dos gestores e administradores possibilitam a escolha de quais indicadores e quando utilizarem, conforme as necessidades, o planejamento das empresas e as áreas analisadas.

De acordo com as áreas, os indicadores podem ser operacionais e apontam seguramente o andamento do trato efetuado em cada processo que seja especializado e executado em determinadas áreas da empresa. Seguem alguns destes principais tipos de indicadores de desempenho:

- **Time to Market:** correspondente ao tempo de lançamento de um produto, iniciando desde a idealização do conceito até a disponibilização pela venda.

- **Lead Time:** consiste no tempo que dura um determinado processo.
- **Stock Out:** indica o número de vezes ou dias em que um item ou produto teve o saldo de estoque zerado.
- **Market Share:** apresenta a “fatia de mercado” conquistada por um produto durante um período.
- **Ociosidade:** calcula o tempo em percentual que uma máquina, equipe ou unidade produtiva fica sem produzir.
- **Giro de Estoque:** relação direta entre consumo, saída e o saldo médio em estoque.
- **Turnover:** taxa de rotatividade de funcionários. É a diferença numérica entre as admissões e as demissões.
- **Tíquete-Médio:** é o total da receita dividido pela quantidade total de vendas.
- **Chargeability:** o principal mecanismo para melhorar a rentabilidade de uma empresa é ter seus funcionários próprios alocados e produzindo, sempre ao máximo possível. É um dos principais indicadores utilizados pela organização. Ele mede o índice de aproveitamento da mão de obra própria da empresa. A meta é estar sempre que possível, o mais próximo de 100%. As variações do índice podem confirmar, por exemplo, a demanda variável e a sazonalidade que uma empresa pode enfrentar ao longo do ano. Os gestores têm uma forte pressão para aumentar o chargeability. No entanto, é importante salientar que ao focar esforços nesse índice, a empresa pode acabar esquecendo da formação de mão de obra, pois, um profissional possui um tempo médio de formação para trabalhar nessa área de cerca de 3 anos.
- **Ebitda:** indicador financeiro muito utilizado por empresas de capital aberto e por analistas de mercado. Conhecido também como Lajida e seu conceito ainda não é claro para muitos. A sigla do inglês é: “*Earning Before Interests, Taxes, Depreciation and Amortization*” que significa o lucro antes dos juros, impostos, depreciação e amortização. O Ebitda é o quanto a empresa, apenas nas atividades de operação geram de recursos, descartando os efeitos financeiros e de impostos. Esse indicador ganha importância porque tradicionalmente a análise do lucro ou prejuízo não tem sido suficiente para avaliar o desempenho real do período, pela influência de fatores difíceis de serem mensurados.
- **Ebit:** conhecido como o lucro na atividade, no que se refere à depreciação e amortização. O Ebit considera estes efeitos contábeis.

Há também outros tipos de indicadores de desempenho utilizados por empresas atuantes exclusivamente em negócios on-line, assim como sites de *e-commerce* ou redes sociais, exemplo:

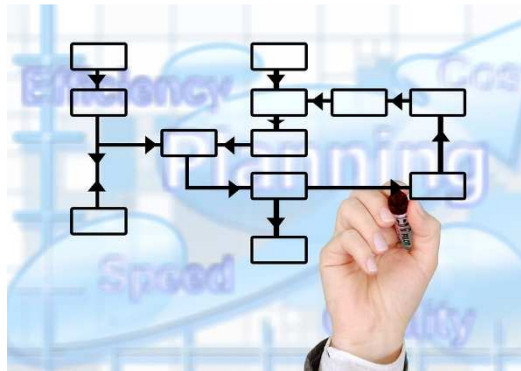
- **Índice de engajamento:** total de visitantes que tiveram algum tipo de interação com a página dividido pelo total absoluto de visitantes.
- **Taxa de Conversão:** total de vendas dividido pelo total de visitas ao site.
- **Taxa de Rejeição:** porcentagem da quantidade de visitantes que acessaram uma única página de um site e saíram, sem navegar em uma próxima página.

- **Índice de influência social:** muito utilizado no Facebook e Twitter, calculado de acordo com o total de postagens curtidas ou retuitadas divididas pelo total geral de postagens publicadas.

Os indicadores de desempenho possuem três valores que auxiliam no cálculo, na avaliação e na tomada de decisão. São eles:

1. **Índices:** são números que apresentam o desempenho obtido em um determinado processo, ou seja, é o que você irá analisar se foi positivo ou negativo.
2. **Metas:** é o valor almejado ou que deve ser “retratado” pelos indicadores para demonstrar o sucesso do processo/etapa.
3. **Tolerância:** o limite de tolerância é um “percentual aceitável” ou não para atingir a meta, avaliando ainda a gravidade por meio dele. Valores fora da faixa de tolerância indicam que o andamento do processo é crítico e deve existir alguma providência (NOVO NEGÓCIO, [2016], on-line).

2.4 Por que utilizar os indicadores de desempenho de processos



Disponível em: <https://pixabay.com/photo-516277/>

Os indicadores basicamente são utilizados para coletar e monitorar as informações de processos, ou seja, de etapas realizadas em uma empresa. Com essas informações, os administradores podem estudar e determinar com cuidado as decisões assertivas, com base em dados consistentes.

Quando uma empresa utiliza indicadores e consegue obter vantagens competitivas. É possível destacar algumas:

- Melhora na assertividade das decisões.
- Coleta, guarda e processamento de informações relevantes.
- Menor tempo gasto na administração e análise de processos.
- Melhora em áreas da empresa, com fácil identificação de falhas.

Além desses destaques, é possível identificar a disponibilização da informação que o administrador necessita a cada etapa do processo. Além disso, proporcionam maior exatidão na tomada de decisão pelo administrador que tem por objetivo trazer mais eficiência e eficácia aos processos.

A maior rapidez e agilidade trazem melhor compreensão e transparência na divulgação dos resultados, com isto, os indicadores de desempenho se tornam

Indicadores de desempenho de processos devem rastrear e seguir o andamento de processos, coletar informações relevantes e disponibilizá-las de forma acessível para que os administradores analisem e tomem as melhores decisões. A eficiência e eficácia trazida aos processos e, conseqüentemente, aos resultados positivos são devido a esses indicadores que são extremamente importantes para a gestão e administração de um processo.

3. Estudo de caso

Métrica é tudo o que pode ser medido, por exemplo: curtidas, fãs, visualizações, comentários e outros. É a quantificação dos dados, úteis para apoiar na noção do que está sendo analisado e o que pode ser concluído a partir desta análise. É também verificação para certificar as metas e as expectativas alcançadas.

Para tratar com métricas é importante alguns cuidados. É necessário destacar bem a compreensão da diferença entre os conceitos de objetivos, métricas, KPIs e estratégias, ou seja:

- 14

Estão apresentados alguns exemplos da diferença dos conceitos:

“Mudar a percepção do consumidor”.....	OBJETIVO
“Fãs em uma página do Facebook”.....	MÉTRICA
“Cliques e alcance”.....	MÉTRICA
“Aumento do número de interações”.....	KPI
“Aumento do tráfego para o site”.....	KPI
“Criar uma promoção no Instagram”.....	ESTRATÉGIA
“Mudar linha editorial do conteúdo”.....	ESTRATÉGIA

Outra questão a ser abordada é a estruturação dos indicadores nos projetos. A ordem com que acontecem os projetos em mídias sociais acontecem exatamente na seguinte ordem:

1. **Briefing:** compreender o negócio, **definir os objetivos, métricas e indicadores de desempenho**. Identificar as redes e as palavras-chave que podem ser monitoradas e definir a estrutura e o escopo dos dados que serão analisados.
2. **Coleta:** configurar as ferramentas, tipos de postagens, foco das palavras-chave estruturadas por assuntos e limitadas no escopo definido.
3. **Classificação:** processar, criar categoria e segmentar os dados conforme a estrutura definida e a expectativa positiva, neutra, negativa.
4. **Análise:** análise dos dados com todas as metodologias necessárias.

É importante que a análise de monitoramento seja feita por meio de um nível macro, permitindo a compreensão de que os níveis externos têm influência direta nos dados das métricas on-line.

Independentemente do meio, mídias digitais ou não, é importantíssimo identificar os conceitos e as atribuições que as informações podem oferecer. Os objetivos, metas e KPIs, respectivamente tratam de: estratégia de monitoramento, relacionamento de números ou porcentagens e avalia os resultados palpáveis do objetivo. Com isto, é observado que as metas estão diretamente ligadas aos KPIs, indicadores de desempenho, que avaliam e identificam se o objetivo traçado foi atingido por meio do que foi planejado.

Fonte: adaptado de Ribeiro (2015, on-line).

4. Considerações finais

Caro (a) aluno (a)!

Os indicadores chave de desempenho - KPIs - têm um papel determinante no desempenho estratégico das organizações.

É claro e notório que mais importante do que uma estratégia na empresa é a sua definição de objetivos claros e aplicáveis. Normalmente, o surgimento dos indicadores ocorre por meio de perguntas simples e básicas, por exemplo: como nossos produtos são vistos pelos nossos clientes?

Para diminuir a margem de erros e problemas que ocorrerão, a gestão por meio de indicadores, reduz o percentual de problemas graves das empresas, independente do seu porte. Em uma organização de grande porte, a implantação da estratégia é de suma importância. É o fato que liga os níveis organizacionais e hierárquicos, desde os diretores até o pessoal de operação. Isso não é sinal de que os empresários de pequeno porte não possam aproveitar os resultados de KPIs para acelerar seus lucros ou ainda reduzir os riscos.

Nos negócios, é surpreendente ver que os KPIs, medidas para determinados objetivos, podem ser simples e inversamente trazer resultados muito mais do que surpreendentes, complexos.

É possível que todo o potencial dos indicadores não tenha ainda sido totalmente dominado, mas cabe aos gestores e administradores conduzirem para novas formas de aproveitar deste método e apoiar as empresas e organizações a alcançar objetivos almejados.

E, por fim, algumas perguntas que auxiliam na reflexão sobre os KPIs e suas aplicabilidades: Está sendo medido o que realmente importa? Os esforços gastos estão nos projetos corretos? Como atingir os objetivos? E para responder a estas questões, devem ser observados tópicos fundamentais para o sucesso das avaliações: os Objetivos, as Iniciativas e os KPIs, pois ao se unirem formam o mais esperado dos resultados, a decisão.

Elementos complementares

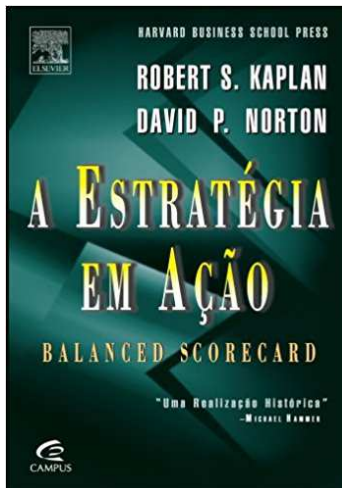


Título: 100 Indicadores da Gestão - Key Performance Indicators

Autor: Jorge Caldeira

Editora: Actual Editora

Sinopse: conheça os 100 indicadores mais utilizados na gestão das empresas. Saiba que tipo de informação dão, como se analisam, onde se vai buscar os dados para o seu cálculo, com que frequência devem ser apurados e como devem ser visualizados graficamente.

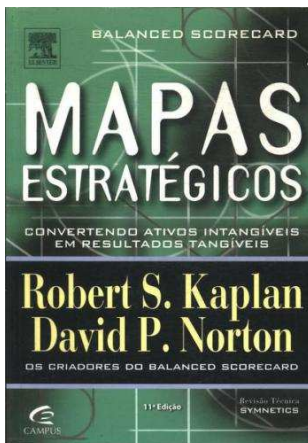


Título: A Estratégia em Ação - Balanced Scorecard

Autores: Robert S. Kaplan e David P. Norton.

Editora: Campus

Sinopse: os autores mostram como altos executivos de diversos setores estão utilizando o Balanced Scorecard para orientar o desempenho atual e focalizar o desempenho futuro. Eles ilustram a utilização de medidas em quatro categorias - desempenho financeiro, conhecimento do cliente, processos internos e aprendizado e crescimento - para alinhar iniciativas individuais, organizacionais e interdepartamentais.



Título: Mapas Estratégicos - Convertendo Ativos Intangíveis em Resultados Tangíveis

Autores: Robert S. Kaplan e David P. Norton.

Editora: Campus

Sinopse: este livro apresenta várias contribuições importantes: um padrão que descreve os componentes básicos de como se cria valor nas perspectivas de processos internos e de aprendizado e crescimento. Temas, baseados nos processos que criam valor, capazes de esclarecer a dinâmica da estratégia e um novo arcabouço para descrever, medir e alinhar os três ativos intangíveis da perspectiva de aprendizado e crescimento - capital humano, capital informacional e capital organizacional - com os processos e objetivos estratégicos da perspectiva dos processos internos.

Referências

KAPLAN, R. S.; NORTON, D. P.; **Using the Balanced Scorecard as a Strategic Management System**. Harvard Business Review. 2007.

MACEDO-SOARES, T. D. L. v. A., RATTON, C. **Medição de desempenho e estratégias orientadas para o cliente: resultados de uma pesquisa de empresas líderes no Brasil**. RAE, v. 39, n. 4, p. 46-59, 1999.

NOVO NEGÓCIO; Disponível em: <<http://www.novonegocio.com.br/lideranca-e-gestao/indicadores-desempenho/>>. Acesso em: 06 jun. 2017.

ORACLE; Disponível em: <<http://www.oracle.com/us/products/applications/cx-metrics-2872398.pdf>>. Acesso em: 05 jun. 2017.

RIBEIRO, H.; **Estudando Social Media - um guia-diário para quem quer trabalhar com marketing digital**, Wordpress; Disponível em: <<https://estudandosocialmedia.wordpress.com/tag/metricas/page/2/>>. Acesso em: 05 jun. 2017.

UNIDADE II - BUSINESS INTELLIGENCE

Introdução

Olá, aluno (a)! Você sabe o que é Business Intelligence (BI) e por que deste assunto? A definição trata de conceitos e métodos para melhorar e auxiliar a tomada de decisões organizacionais usando fatos que são gerados a partir de informações e sistemas existentes.

A importância dos sistemas de BI está na sua utilização e no apoio das organizações em suas decisões, considerando uma série de benefícios, como: maior rentabilidade, menores custos, aumento da eficiência, aumento do valor de mercado, maior satisfação dos colaboradores, além, dos melhores resultados alcançados, que é um dos maiores objetivos e desafios das organizações.

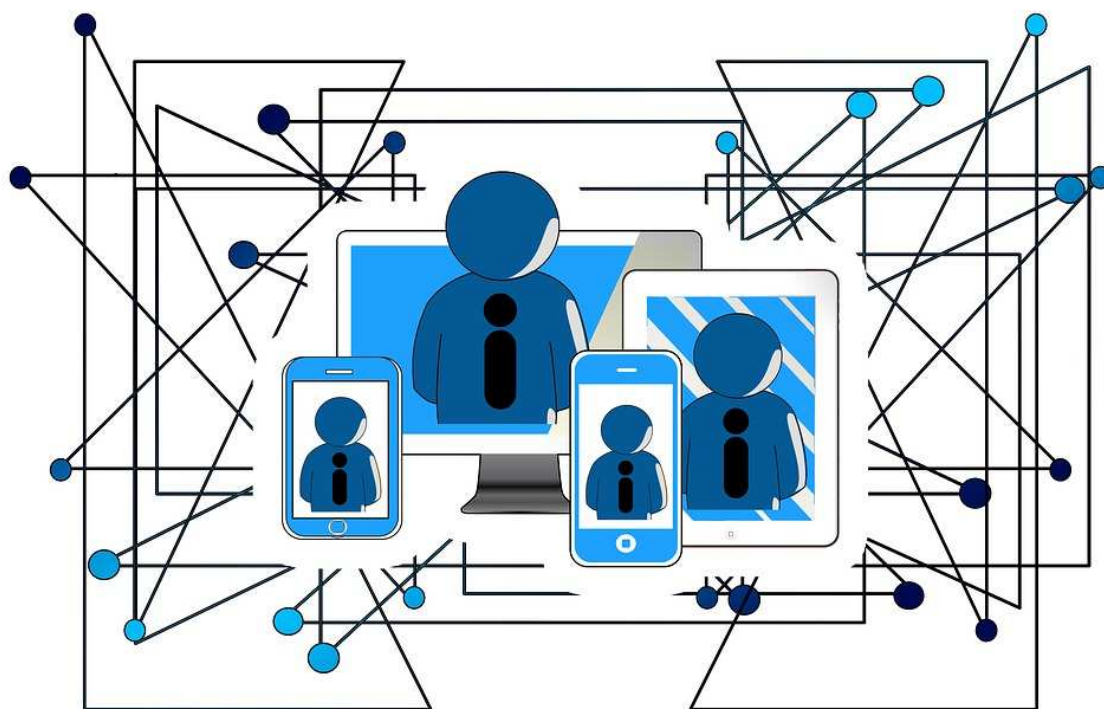
Os processos de BI devem ser elaborados sobre bases de dados separadas, e não na base de transações, onde ocorrem as operações cotidianas, para não paralisar servidores e atrapalhar as operações diárias. Imagine uma empresa brasileira com franquias espalhadas por todo o Brasil e de repente os servidores estão em alto processamento e os compradores da empresa não conseguem realizar nenhuma compra, pois algum executivo está realizando análises complexas.

A solução é gerar uma só base de análise, chamada de base *On-Line Analytical Processing* (OLAP). Com este conceito, identificamos um Data Warehouse (DW) que é uma Base de Dados que centraliza a composição dos dados copiados de outras bases, chamadas *On-Line Transactional Processing* (OLTP).

Outra ligação importante com o BI são as volumosas bases de dados contemporâneas, que atraem dados de todos os tipos e em um crescimento exponencial, a chamada Big Data. As análises sobre essas Bases de Dados devem ser realizadas por sistemas capazes de efetuar esta atividade, privilegiando a performance e a consistência dos dados, sendo o BI a ferramenta mais indicada. Diante dessa abordagem, a partir de agora você está convidado a conhecer um pouco mais o tema proposto nesta breve introdução.

Então, vamos lá! Ótimo estudo!

1. Conceito de Business Intelligence (BI)



Disponível em: <https://pixabay.com/photo-2496194/>

Há muitas definições para Business Intelligence (BI) e em vários casos alguns profissionais afirmam estarem trabalhando com BI quando estão, na verdade, apenas gerando informações com sistemas de informações gerenciais, ou *Enterprise Resource Planning* (ERP), utilizando ferramentas para geração de relatórios e convertendo em análises visuais como gráficos e *dashboards*.

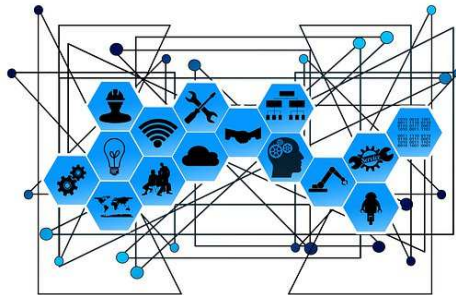
O objetivo de um processo de BI é encontrar causas, explicações e padrões, considerando o volume de dados cada vez mais complexos, além do uso de ferramentas mais avançadas, principalmente no que se refere ao software. Por isto, existe o uso do termo Cientista de Dados, que é tão atual e necessário para este trabalho.

É importante enfatizar que os dados são importantes e necessários para o processo, incluindo a forma e as condições como são coletados e armazenados. Não basta abordar estes trabalhos de coleta, armazenamento e processamento, é necessário discutir formas dinâmicas de análise e exploração dos resultados, considerando que os dados analisados têm qualidade.

O BI surgiu há milhares de anos com os povos Persas, Egípcios e outros povos. Eles, naquele tempo, antecipavam o cruzamento de informações provenientes da natureza, assim como períodos de chuvas e comportamento dos mares para diversas decisões. No livro “A Arte da Guerra”, Sun Tzu (2002) relata a necessidade de deter todo o conhecimento de suas fraquezas e virtudes, além de todo o conhecimento das fraquezas e virtudes do inimigo. A falta desse conhecimento pode resultar na derrota.

Simplificando, é possível descrever BI como: o ato de transformar dados em informações úteis. O termo *Business Intelligence*, ou Inteligência de Negócios em português, em uma visão resumida, significa coletar, organizar, analisar e compartilhar os dados de diversas fontes, com executivos de maior importância na empresa. As informações relevantes são transformadas em importantes decisões para o futuro da empresa.

1.1.A Evolução da Informação



Disponível em: <https://pixabay.com/photo-2496193/>

A Tecnologia da Informação com computadores, redes de comunicação, softwares e diversos processamentos teve nas organizações um importante papel para dar entrada, processar e armazenar volumosa quantidade de dados e ainda apoiar em cálculos e operações de resultado. Dessa maneira, os primeiros processos controlavam estoque, folha de pagamento e contabilidade entre outros. Esses sistemas são de processamento de transações e têm característica rotineira.

Com o uso desses sistemas foi possível identificar que desses resultados primários era possível submetê-los a novos processamentos e extrair novas informações e gerar outros relatórios. Ou seja, do uso de um controle de estoque, começou ser possível saber e refletir sobre quais os produtos estavam sendo mais vendidos, os que menos saíam e gerar um gráfico das saídas médias. Igualmente em um sistema de folha de pagamento foi identificado o cargo ou setor que tinham mais custos e relacionar ao número de funcionários e diversas outras análises. Na contabilidade, da mesma forma, processar a relação entre gastos e expectativa de receitas.

Os relatórios avançaram para inovadores Sistemas de Informações Gerenciais, conhecidos pelo acrônimo SIGs. Foram inseridos modelos de gráficos e painéis com a apresentação de diferentes informações, conhecidos como *dashboards*. Bertin (1983) apresenta e descreve tipos de gráficos diferentes e suas aplicações.

Os sistemas de informações gerenciais são muito importantes e têm uso contemporâneo em qualquer empresa. Os administradores tomadores de decisão, gestores e executivos necessitam de informações rápidas e que auxiliem a tomada de decisão.

Anteriormente, para encontrar características de forma a ajudar a compreender o que estava registrado ou processado, eram necessários os modelos descritivos, que até os dias atuais buscam identificar os padrões. O BI inicia neste ponto, no entendimento do porquê das coisas e como ocorrem, formando contextos das causas ou explicações para determinados eventos ou situações.

O processo pode ser dividido como um todo, levando em consideração o que se quer saber em relação a como encontrar suas respostas, conforme apresentado no Quadro 1.

Quadro 1 - Relação das questões a serem atendidas e suas análises

Pergunta	Exemplo	Análise
O que aconteceu?	Totais de venda no mês anterior.	Existem os SIGs, que buscam informações em sistemas transacionais e geram relatórios.
O que está acontecendo?	As vendas estão aumentando ou reduzindo?	É possível também utilizar SIGs ou sistemas de Data Mining para encontrar padrões estatísticos nos dados.
Por quê?	Por que as vendas estão caindo?	Aqui inicia o BI, investiga os motivos para os eventos observados.
O que acontecerá no futuro?	Ao manter os níveis de venda e diminuir o preço de venda, o que acontecerá com o lucro?	As previsões e análises com base em simulação de variáveis e condições, uso de Business Analytics e Sistemas de Apoio à Decisão.
O que gostaríamos que acontecesse?	Expectativa de aumento da receita em 10%.	Essencialmente planejamento e definição de metas. É necessário compreender as causas e as variáveis interrelacionadas.

O que aconteceu?	Totais de venda no mês anterior.	Existem os SIGs, que buscam informações em sistemas transacionais e geram relatórios.
O que está acontecendo?	As vendas estão aumentando ou reduzindo?	É possível também utilizar SIGs ou sistemas de Data Mining para encontrar padrões estatísticos nos dados.
Por quê?	Por que as vendas estão caindo?	Aqui inicia o BI, investiga os motivos para os eventos observados.
O que acontecerá no futuro?	Ao manter os níveis de venda e diminuir o preço de venda, o que acontecerá com o lucro?	As previsões e análises com base em simulação de variáveis e condições, uso de Business Analytics e Sistemas de Apoio à Decisão.
O que gostaríamos que acontecesse?	Expectativa de aumento da receita em 10%.	Essencialmente planejamento e definição de metas. É necessário compreender as causas e as variáveis interrelacionadas.

Fonte: o autor.

1.2. B.I. X Sistema de Informação Gerencial



Disponível em: <https://pixabay.com/photo-2302728/>

O BI é usualmente relacionado aos softwares que geram relatórios, os conhecidos Sistemas de Informações Gerenciais (SIGs) ou *Management Information Systems* (MIS). Eles geram relatórios ou gráficos, de forma sintética ou em comparações.

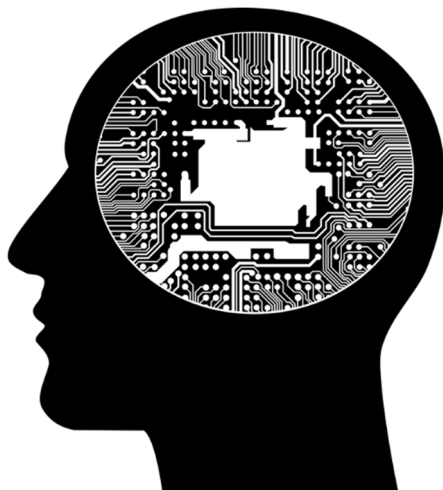
O BI gera informações novas, processadas e que não eram percebidas nos dados, ou ainda permite a visualização das informações com a percepção rápida e fácil de algo novo. Os relatórios que apontam os produtos mais lucrativos ou vendidos, melhores vendedores, lojas mais rentáveis, sazonalidade dos produtos, entre outros, são exemplos claros.

Os sistemas são muito importantes para as empresas. E para somar, BI deve ser utilizado com mais detalhes. O processo de BI ajuda no descobrimento das causas desses acontecimentos ou descobertas. Com isto, a associação é simples, o SIG aponta o produto mais vendido e o BI procura o porquê deste produto ser o mais vendido ou porque os outros não vendem mais. O SIG apresenta o momento de maior venda de um produto, o BI busca o porquê do produto vender mais neste momento e menos em outros.

Os SIGs auxiliam na compreensão do que aconteceu ou do que está acontecendo, assim como na descoberta da taxa de crescimento de nossas vendas e totais de venda no mês anterior. O BI já procura por causas e explicações, assim como no porquê das vendas estarem caindo.

Tanto SIG quanto BI são tipos de sistemas que procuram auxiliar na tomada de decisão, conjugando o objetivo geral dos sistemas de informação. Contudo, o método que o sistema apoia é que pode ser diferenciado para cada tipo de sistema.

1.3. Dados, Informação, Conhecimento e Inteligência



Disponível em: <https://pixabay.com/photo-2099119/>

É importante compreender os conceitos aplicados à construção da inteligência. A evolução passa dos dados, que ao serem processados geram informação, que por sua vez geram conhecimento e transcende esta evolução com a inteligência, conforme apresentado na Figura 01.

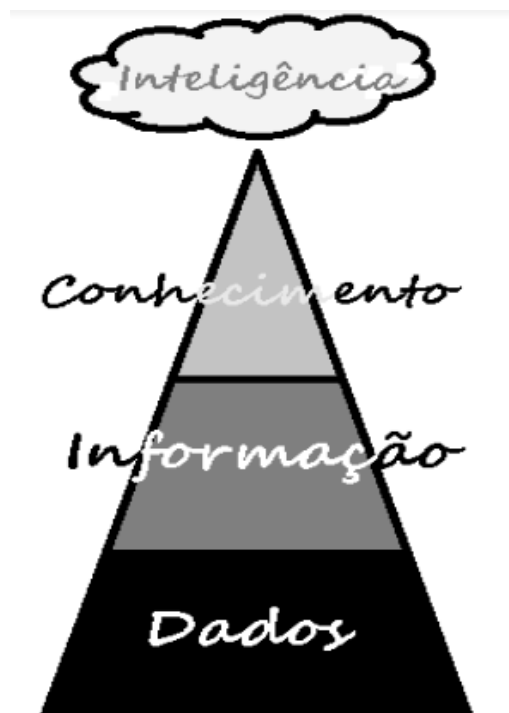


Figura 01 - Formação da inteligência
Fonte: o autor.

A Tabela 1 apresenta o valor 43 na primeira linha com a segunda coluna de dados. Os dados são representações de informações. Sozinhos são indiferentes e não contribuem. Porém, quando existe a compreensão de que o número 43 significa a idade da pessoa Natal, em anos, este dado está sendo transformado em informação. As pessoas necessitam de informações para trabalhar, mas o mundo computacional armazena e processa dados.

Tabela 1 - Dado, Informação e Conhecimento

Pessoa	Idade	Saldo Médio	Cidade
Natal	43	10000	Ribeirão Preto
José	40	5000	Curitiba
João	39	9300	Porto Alegre
Silva	35	4000	Brasília
Regina	44	10500	Ribeirão Preto

Fonte: o autor.

Neste mesmo exemplo, é possível observar que todos os clientes da cidade de Ribeirão Preto têm saldo médio maior que 10 mil. É possível observar que isto não está destacado ou estabelecido na tabela. O processamento destas informações é o conhecimento. Só é possível chegar a este conhecimento processando as informações diferentes e confrontando os seus resultados. Conhecimento é formado por meio das informações, mas está acima. As pessoas são alimentadas de muitas informações no dia a dia, porém apenas

uma parte fica retida, outro detalhe é que nem tudo é útil e nem tudo precisa ser aproveitado. Contudo, o que é utilizável forma o conhecimento das pessoas.

A Inteligência ou Sabedoria está acima de conhecimento. Em um exemplo hipotético, um grupo de pessoas é colocado em um ambiente fechado, a tarefa dada a estas pessoas seria de fazer algo que utilize a força bruta, e que poderia machucar uns aos outros. Nesta hipótese, todos possuem a mesma força física e foram educados tradicionalmente em suas famílias e estudaram em escolas igualmente tradicionais. Ou seja, possuem o mesmo nível de conhecimento, cultura, espiritualidade, entre outros. Algumas dessas pessoas resolveram o problema e outras não. A questão que fica é, porque isso acontece, sendo que a força física e os conhecimentos são os mesmos? A inteligência é a diferença que está na forma e na conduta como cada um utiliza o conhecimento e suas conexões no cérebro. Ou seja, inteligência é saber resolver problemas utilizando o conhecimento que possui e pode se estender a resolver novos problemas com analogias e adaptações.

BI, então, como o nome "inteligência" indica, deve ajudar as pessoas e as organizações a resolverem seus problemas e alcançarem seus objetivos. Com isto, o próprio nome já relaciona *Business Intelligence* à inteligência que ajuda pessoas e instituições a alcançarem suas metas e objetivos por meio da resolução de seus problemas.

1.4. Objetivos do BI



Disponível em: <https://pixabay.com/photo-1263422/>

Primeiramente, é importante expressar que BI é um processo. Existem diversas técnicas, tecnologias, ferramentas e softwares para BI, contudo BI é um processo que tem relação com:

- Métodos.
- Técnicas.
- Tecnologias.
- Pessoas.
- Informações.
- Fontes de informações.
- Métricas.
- Ferramentas, entre outros.

O processo de BI objetiva a busca das causas e explicações para situações, eventos e resultados com ou sem expressividade. Os resultados podem ser positivos ou negativos, pois para o BI é indiferente, o que deve estar em evidência é a localização das causas dos problemas e as melhores práticas de sucesso.

Não é o suficiente ter o conhecimento do problema mais comum nas máquinas de uma indústria, a empresa necessita conhecer o porquê disto. Com esta questão respondida, é possível atuar sobre as causas e minimizar os efeitos negativos. Não basta saber qual o melhor produto, a empresa precisa saber por que ele é o melhor, para que as características deste produto possam ser transmitidas para os demais da linha de produção.

O BI pode integrar e utilizar os dados dos sistemas gerenciais, de Ferramentas de Data Mining e também sistemas Enterprise Resource Planning (ERP). O BI está na ponta de todo o processo de transformação da informação, para que as análises sejam rápidas e disponíveis, com isso está na proximidade dos tomadores de decisão.

O maior objetivo do BI é terminar com a indecisão nos processos. As análises fazem com que o tratamento estatístico seja uma ferramenta alinhada com as massas de dados disponíveis e subutilizadas.

O conhecimento nos torna mais inteligentes ou com maiores capacidades de decisão. O BI também busca encontrar explicações em dados, para que eventos sejam fundamentados em dados. Não basta saber o que está ocorrendo, é necessário analisar as causas para que o sucesso seja repetido ou o fracasso evitado.

Um dos objetivos do BI também é a busca por padrões. Na massa gigantesca de dados, é necessário encontrar uma ordem e padrão para que os dados possam ter o sentido e a utilidade almejada. Uma base de clientes em que não é possível identificar quem são os clientes, o que eles têm de produtos adquiridos, quais suas características empresariais, não é útil para nada além de desviar a atenção de um departamento comercial.

1.5. Por dentro dos Fatos



Disponível em: <https://pixabay.com/photo-1093891/>

Qual a relação existente entre Data Warehouse, Business Intelligence e Big Data? A interligação e a perspectiva desses três conceitos são cada vez mais utilizados pelas maiores empresas do mundo. As análises geradas por essas tecnologias geram uma visão mais analítica em qualquer negócio e auxiliam diretamente nas melhores tomadas de decisões. Os três conceitos levam em consideração o volume gigantesco de informações em diversos formatos e que contribuem completamente, pois mesmo que diferenciados eles têm em comum uma nova forma de trabalhar e extrair dados, seja de forma estruturada ou não, e que influencia diretamente na tomada de decisões estratégicas. Talvez este seja o ponto comum das tecnologias Data Warehouse (DW), Business Intelligence (BI) e Big Data. Afinal, o objetivo de qualquer um desses conceitos é representar um resultado dentro do diferencial competitivo das corporações, mas a forma com que são utilizados é que faz toda a diferença.

Fonte: o autor.

2. Data Warehouse (DW)



Disponível em: <https://pixabay.com/photo-286386/>

Um Data Warehouse(DW), também conhecido como “Armazém de dados”, é um Banco de Dados contendo dados extraídos do ambiente de produção (Sistema que alimenta as bases de dados) de uma empresa. Esses dados são determinados e depurados com uma otimização própria para o processamento e apresentação de resultados e não para processamento de transações. Em geral, um DW permite a consolidação de recursos de dados, além de Banco de Dados relacionais podem ser extraídas informações de planilhas eletrônicas, textos, entre outros.

Na grande maioria dos casos, os dados são provenientes de diversas fontes independentes, podendo ser internas da empresa ou externas para a composição e consolidação destes dados.

Em geral, os processos de BI devem ser feitos e processados sobre bases de dados separadas das bases transacionais, onde as operações do dia a dia

ocorrem. Isto é importante para não onerar os servidores ou atrapalhar as operações dos colaboradores com lentidão nos processos.

Em um exemplo fica mais fácil de imaginar. Pensando em uma empresa com unidades em todo o mundo, e em determinado momento seus colaboradores não conseguem realizar nenhuma operação no sistema, isto porque os servidores de aplicação e Banco de Dados estão lentos, ocupados com algum executivo realizando análises complexas de dados.

Então, a solução para isto é gerar uma base só para as análises necessárias, e também somente com os dados necessários, pois, muitas vezes não são todas as tabelas e atributos que são necessários para realizar estas análises. Essa base é chamada de *On-Line Analytical Processing* (OLAP).

Esse é o conceito de um DW. A centralização de bases formadas por dados copiados de diversas bases chamadas *On-Line Transactional Processing* (OLTP). Com isto, são separadas as bases de dados e os servidores, sendo um esquema montado para as aplicações transacionais em nível operacional da empresa que contemplam as tarefas de inclusão, exclusão, alteração e consulta simples de dados e valores, e outro esquema com dados só para análise, sendo dados não voláteis, somente incluídos, e que apoiam as tomadas de decisões táticas e estratégicas.

2.1. Diferenças entre OLTP e OLAP



Disponível em: <https://pixabay.com/photo-767776/>

O OLTP armazena no Banco de Dados todas as transações de negócios registradas em um sistema. Esse tipo de base de dados funciona bem em sistemas que gravam pequenas transações realizadas em tempo real, que ocorrem frequentemente e de forma rápida e instantânea. Os dados ficam à disposição para alterações e modificações. É uma base que necessita de backup regularmente para a garantia de continuidade e disponibilidade.

O OLAP é um pouco diferente, tornando-se outra alternativa. Utilizado na tomada de decisões, proporciona uma visão dos dados orientada à análise, além de uma navegação rápida e flexível. O OLAP recebe dados do OLTP para que possa realizar as análises e essa carga de dados acontece conforme a necessidade da empresa. Sendo um sistema para tomada de decisões, que não realiza transações, pois sua finalidade são as consultas. Possui dados atuais e históricos e não há necessidade de backups regularmente.

É importante compreender a diferenciação destes dois tipos de informações, para que a utilização e extração das informações sejam devidamente conduzidas conforme suas disponibilidades e importâncias.

OLAP é uma tecnologia que possibilita aos gerentes e executivos extraírem de forma consistente e rápida, várias possibilidades de visões da informação, ou seja, facilita a análise multidimensional da informação.

3. Projeto de Business Intelligence (BI)



Disponível em: <https://pixabay.com/photo-2394312/>

Projetos de BI tendem a ser complexos. Para simplificar a compreensão, o projeto de BI pode ser dividido em treze etapas.

3.1. Etapas de um projeto de BI

Levantamento dos Key Performance Indicator (KPIs) e definição do escopo:

- Estar acessível à necessidade de informações dos administradores e às entregas e atualização rotineiras.
- Verificar se há algum relatório sendo entregue.
- Entender a forma de entrega e quem será o responsável.
- Se já houver um relatório, verificar quem é o responsável pela entrega e quais dados são mais úteis.
- Identificar os usuários do sistema.
- Compreender as necessidades do usuário e ajustar os processos de utilização.

Listar dados para confeccionar o plano:

- Cruzar dados essenciais para confecção dos KPIs.

Entrevistar DBA:

- Apresentar os dados necessários para a demanda e ajustar todo o necessário no Banco de Dados.

Analisar dicionário de dados existente:

- Será importante para identificação e uso dos dados.

Entrevistar o departamento que utilizará:

- Buscar as informações nas áreas envolvidas no processo de entrada de dados e de processos que estejam envolvidos.
- Compreender como o usuário visualiza os dados.

Entrevistar a área de Infraestrutura:

- Apresentar a estrutura da informação requisitada e apresentar a execução e atualização demandada.
- Ver tecnicamente a viabilidade da implantação.
- Se não for viável, retornar à primeira etapa e apresentar as limitações aos administradores.

Criar a Query:

- Solicitar ao DBA a criação da estrutura de dados no banco.

Instalar o BI:

- Instalar o programa de BI.

Apresentar propostas de Extração Transformação e Leitura (ETLs) e Cubos:

- Criar os ETLs e Cubos.
- Criar relatório para a primeira validação.
- Caso reprovado, retornar à etapa

Criar os Data Marts:

- Criar e testar a rotina.

Criar os Dashboards:

- Caso reprovado, voltar a etapa.
- Testar a atualização.
- Caso não dê certo, verificar o ponto de retorno.

Implantar:

- Gerar os relatórios finais e disponibilizar aos administradores.

Entregar a documentação:

Entregar o relatório de implantação do BI.

3.2. Pentaho BI Suíte

Proveniente de empresas como Business Objects, Cognos, Hyperion, JBoss, Oracle, RedHat e SAS, a Pentaho Corporation, empresa norte-americana, tornou-se referência em Inteligência de Negócios de código aberto, fundada em 2004 por uma equipe de veteranos da indústria de Inteligência de Negócios e vendida para a Hitachi em 2015.

A Pentaho possui, até então, dois modelos de comercialização. Um modelo com licença anual, que dá direito a utilização de suporte e o outro modelo, sem pagar licença de uso, utiliza a versão community onde não existe suporte, nem custo anual, e o suporte é feito pela comunidade de desenvolvedores e entusiastas Pentaho espalhados por todo o mundo.

A solução completa do BI Pentaho utiliza diversos componentes. A coletânea desses componentes é chamada de Pentaho BI Suite, ou seja, um conjunto completo de programas utilizados para uma solução de BI.

Com a suíte é possível realizar a integração de dados, construir processos de ETL visualmente, criar cubos OLAP, construir consultas e relatórios com filtros, com fórmulas, com sub-relatórios, com alertas, relatórios ad-hoc, análises interativas, painéis, mineração de dados entre outras opções.

Essas opções tornam a Suíte BI Pentaho muito popular no mundo *open source*.

A aplicação Pentaho BI Server é responsável pela gerência dos indicadores, o compartilhamento entre os usuários, o controle de acesso, entre outras opções. Algumas das tarefas:

- Auditoria.
- Autenticação de usuários.
- Integração com mineração de dados.
- Integração com painéis.
- Integração com relatórios.
- Integração de análises.
- Logs.
- Motores de regras de negócio.
- Serviços web.

3.2.1. O download Pentaho

A melhor forma de efetuar o download do Pentaho é baixando os aplicativos que serão úteis em primeiro lugar, inclusive para os testes. É necessário que as estruturas de pastas sejam definidas. Para executar e testar a suíte Pentaho não há necessidade de instalar aplicativos. Para tanto, o Pentaho pode ser copiado na pasta que será utilizada e a partir disso, ele estará operacional. Algumas configurações de ambiente podem ser necessárias.



Figura 02 - Download do Pentaho
Fonte: o autor.

A estrutura para baixar é a seguinte:

- Files.
- Escolher aplicativos.
 - Business Intelligence Server contém:
 - Pentaho UserController (PUC).
 - Pentaho Administration Console (PAC).
- Data Integration
 - Pentaho Data Integration (PDI).
 - Extrair em um diretório seu, em um lugar só.
- Start-pentaho.bat
 - Pode não executar o Java ARchive (JAR).
 - SOLUÇÃO: Configurar variáveis de ambiente.

3.2.2. Instalação e configuração do Pentaho

Para instalar o Pentaho e as ferramentas e plugins citados acima, é necessário apenas a descompactação dos arquivos.

3.2.3. Criação da pasta

Após fazer o download dos arquivos necessários da suíte, copie os arquivos compactados para a pasta que deseja instalar o Pentaho, deve ser um local de sua preferência, e descompacte os arquivos.

3.2.4. Instalação e configuração do Java

Verifique se a versão do java instalado em sua máquina é compatível com a versão Pentaho que está instalando. Certifique que as variáveis de ambiente do Java estejam configuradas no computador. Para verificar:

Clique com o botão direito em Meu computador -> Propriedades -> Configurações Avançadas -> Variáveis de ambiente -> Variáveis do sistema -> Adicionar ou Alterar:

- **JAVA_HOME** : Diretório do seu Java: No meu caso > C:\Program Files\Java\{versão do seu java 7}\
- **PENTAHO_JAVA_**: Diretório do seu Java RuntimeEnvironment (JRE): No meu caso > C:\Program Files\Java\{versão do seu java 7}\JRE
- **Variável PATH** -> Se não possuir a variável PATH, deve criar e adicionar o código do diretório do Bin do seu Java, por exemplo: C:\Program Files\Java\{versão do seu java 7}\bin, e se acaso já existir, deverá concatenar ao final da linha o diretório separado

com ponto e vírgula, variando de acordo com o seu sistema operacional. Exemplo imagem na Figura 03.



Figura 03 - Exemplo de edição do path de ambiente.
Fonte: o autor.

É importante que utilize o Java 7, pois somente ele é compatível com a ferramenta tomcat que vem junto com o Pentaho.

- Crie uma pasta chamada Pentaho em seu c:
- Extraia os arquivos baixados para dentro desta pasta criada.
- É importante que todos os dados, de PUC e PDI fiquem dentro da mesma pasta.

4. Configurações diversas



Disponível em: <https://pixabay.com/photo-769300/>

O PUC funciona como padrão na porta 8080, porém, essa porta é bastante utilizada, então se acaso precisar mudar, o que é recomendável, será preciso modificar dois arquivos:

- No primeiro `<\pentaho\biserver-ce\tomcat\conf\server.xml>`, conforme Figura 04, alterar a porta em: `<ConnectorURIEncoding="UTF-8" port="8080" protocol="HTTP/1.1">`;

```

    APR (HTTP/1.1) connector: /docs/8080.html
    Define a non-SSL HTTP/1.1 Connector on port 8080
-->
<Connector URIEncoding="UTF-8" port="8080" protocol="HTTP/1.1"
    connectionTimeout="20000"
    redirectPort="8443" />
<!-- A "Connector" using the shared thread pool-->
<!--
<Connector URIEncoding="UTF-8" executor="tomcatThreadPool"
    port="8080" protocol="HTTP/1.1"
    connectionTimeout="20000"
    redirectPort="8443" />

```

Figura 04 - Porta para conexão do Pentaho
Fonte: o autor.

- No segundo, editar o arquivo web.xml no seguinte caminho: <\pentaho\biserver-ce\tomcat\webapps\pentaho\WEB-INF\>. Nesse arquivo, conforme Figura 05, deverão ser alteradas todas as referências à porta 8080. Uma sugestão seria alterar para a porta 8045;

```

<context-param>
    <param-name>fully-qualified-server-url</param-name>
    <param-value>http://localhost:8080/pentaho/</param-value>
</context-param>

<context-param>
    <param-name>locale-language</param-name>
    <!-- for example 'fr' for French, 'de' for German, da for
    <param-value></param-value>

```

Figura 05 - Porta para conexão do Pentaho
Fonte: o autor.

- Após as alterações é necessário reiniciar o biserver-ce, para reiniciar:
 - Se **Windows**: primeiro: <\pentaho\biserver-ce\stop-pentaho.bat> depois: <\pentaho\biserver-ce\start-pentaho.bat.>.
 - Se **Linux**: primeiro: <./pentaho/biserver-ce/stop-pentaho.sh> depois: <./pentaho/biserver-ce/start-pentaho.sh.>.
- Depois será possível logar no portal e testar a alteração, no exemplo da alteração será: <http://localhost:8045/pentaho>.

5. Inicialização das ferramentas



Disponível em: <https://pixabay.com/photo-593312/>

5.1. Pentaho Administration Console (PAC)

- Abrir o arquivo: <\pentaho\biserver-ce\start-pentaho.bat>.
- Abrir o arquivo: <\pentaho\administration-console\start-pac.bat>.
- Acessar: <localhost:8099> ou porta utilizada.
- Padrão: Login: admin / Senha: password

5.2. Pentaho UserController (PUC)

- Abrir o arquivo: <\pentaho\biserver-ce\start-pentaho.bat>.
- Abrir o arquivo: <\pentaho\administration-console\start-pac.bat>.
- Acessar: localhost:8080 (ou porta definida nas configurações)
- Padrão: Login: joe / Senha: password

5.3. Pentaho Data Integration (PDI) – KETTLE

- Copiar o arquivo de conexão (mysql-connector-java-5.1.17.jar) MySQL para a pasta JDBC da ferramenta para fazer possibilitar a conexão ao banco MySQL.
- Abrir o arquivo: <\pentaho\biclient\data-integration\spoon.bat>.

5.4. Pentaho SchemaWorkBench (PSW) – MONDRIAN

- Copiar o arquivo de conexão (mysql-connector-java-5.1.17.jar) MySQL para a pasta JDBC da ferramenta para fazer possibilitar a conexão ao banco MySQL.
- Abrir o arquivo \pentaho\biclient\schema-workbench\workbench.bat.

5.5. Pentaho Metadata Editor (PME)

- Abrir o arquivo: <\pentaho\biclient\pme-ce-4.8.0-stable\metadata-editor\metadata-editor.bat>.
- Configurar permissões: Tools -> Security -> URL: <http://localhost:9090/pentaho/ServiceAction> – Usuário: Joe / Senha: password

5.6. Pentaho Report Designer (PRD)

- Abrir o arquivo: <\pentaho\biclient\prd-ce-3.9.1-GA\report-designer\report-designer.bat>.

5.7. CTOOLS

- Acessar localhost:9090 (ou porta definida) e clicar no ícone do Ctools.

5.8. SAIKU

- Acessar localhost:9090 (ou porta definida) e clicar no ícone do Saiku.
- Com isto o ambiente servidor estará configurado para executar o Pentaho.

5.9. Utilização Pentaho



Disponível em: <https://pixabay.com/photo-1839876/>

O Pentaho BI possui uma instalação pré-configurada para demonstração do uso de relatórios, cubos e dashboards da base de dados Steel Wheels. Para executar esta simulação e também utilizar o BI, é necessário iniciar o servidor, veja os procedimentos de iniciação do servidor.

5.9.1 Iniciar o BI Server

Para iniciar o Servidor de BI, seguir os seguintes passos:

- No windows utilize o arquivo: <\pentaho\bi-server-ce\start-pentaho.bat>.
- No Linux utilize o arquivo: </pentaho/bi-server-ce/sh/start-pentaho.sh>.
- Se estiver utilizando a porta padrão, acesse o BI Server por meio da URL: <http://localhost:8080/pentaho>.
- Os usuários de testes estão apresentados na própria tela de login. Pode entrar com o usuário “joe” para navegar na aplicação.
- Todas as simulações de relatórios, cubos e dashboards existentes, podem ser visualizados da base de dados *Steel Wheels*.

5.9.2 Utilizar Pentaho Data Integration (PDI)

De todos os aplicativos da suíte do Pentaho, o PDI é uma das ferramentas mais importantes, e pode ser utilizada com qualquer outra aplicação e finalidade. Inclusive, o PDI é muito utilizado para transmissão de dados entre portais e ainda para importação e exportação de dados.

Com sua forma simples de extrair ou integralizar dados e informações, é uma das ferramentas mais importantes de todo o processo. As três etapas podem ser observadas na Tabela 02.

Tabela 02 - Etapas de ETL

Extração	Dados de diferentes fontes e formatos. Identificação de mudanças desde a última extração.	Validação e descarte de dados de acordo com regras e padrões
Transformação	Dados de acordo com requisitos técnicos e de negócio	Conversão dos tipos de dados, filtragem de dados e sumarização
Carregamento	Dados transformados em uma base de dados	Reescrita dos dados e adição de novas informações

Fonte: o autor.

A transformação pode utilizar qualquer entrada de dados que esteja estruturada, assim como: planilhas eletrônicas, arquivos txt, arquivos XML, os mais diversos Bancos de Dados existentes no mercado, incluindo NoSql.

As atividades de transformação são:

- Validação dos dados
 - Verificação se os dados estão corretos e precisos.
 - Filtragem de dados inválidos.
- Limpeza dos dados
 - Correção de dados inválidos.
- Decodificação
 - Conversão de atributos (numéricos, categóricos) para adequação a um padrão ou regra.

- Agregação
- Geração e gerenciamento de chaves
 - Dimensões identificadas por chaves substitutas.

As atividades de carregamento são:

- Carregamento das tabelas de fatos
 - Adição de linhas à tabela de fatos.
 - Atualização de atributos de status.
- Carregamento e manutenção das tabelas de dimensões
 - Adição e atualização de linhas das tabelas de dimensões.

Para utilizar o PDI, executar spoon.bat ou Kettle.exe no Windows, ou ainda spoon.sh no Linux.

Dicas de configuração da área de trabalho do Spoon em: Menu Editar/Opções:

- Na aba “General”
 - Show tipsat startup?
 - Show welcomepageat startup?
 - ...
- Aba “Look-and-feel”.
 - Preferredlanguage.
 - ...
- As mudanças estarão visíveis após reiniciar o Spoon.

O PDI trabalha com dois tipos básicos de componentes: transformações e jobs. As características de transformações e jobs, além de definirem o fluxo do processo de ETL, também contém os metadados do processo de ETL: Descrição dos dados, Fontes de entrada e saída, Scheduling e Scripting. Uma transformação ou job consiste de uma coleção de itens interconectados.

As conexões entre os itens das transformações e jobs são chamadas de Hop's e são as ligações, *pipeline*, do fluxo de registros.

As Transformações consistem em uma coleção de steps (passos) de transformação, onde cada step denota e executa uma operação do processo de ETL. A saída de um step produz um conjunto de registros e os fluxos dos steps da transformação ocorrem de forma simultânea e assíncrona e geram um arquivo .ktr.

Os Jobs consistem em uma coleção de transformações ou de steps de jobs, onde cada entrada do job denota e executa uma tarefa do processo de ETL. A saída de cada entrada do job produz um status de execução, e os fluxos dos steps do job ocorrem de forma sequencial e geram um arquivo .kjb.

O metamodelo dos componentes do PDI pode ser observado na Figura 06.

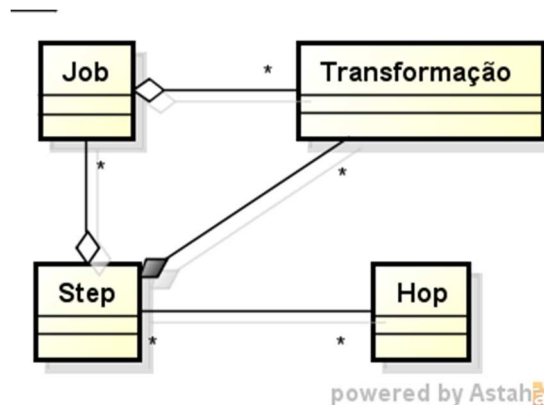


Figura 06 - Metamodelo dos componentes do PDI
Fonte: o autor.

Outros componentes do PDI são os Repositórios que possuem os metadados das transformações e dos jobs que podem ser persistidos em um Banco de Dados. Também as Ferramentas, que podem ser:

- **Spoon:** Ferramenta para desenvolvimento visual.
- **Pan:** executa as transformações em linha de comando.
- **Kitchen:** executar os jobs em linha de comando.
- **Carte:** servidor para execução remota de transformações e jobs.

5.10 Resultados de um B.I.

Os resultados podem ser apresentados de diversas formas.

Os relatórios, conforme Figura 07 são apresentações estáticas em que podem ser incluídos filtros e diversos gráficos e informações em um mesmo processamento.

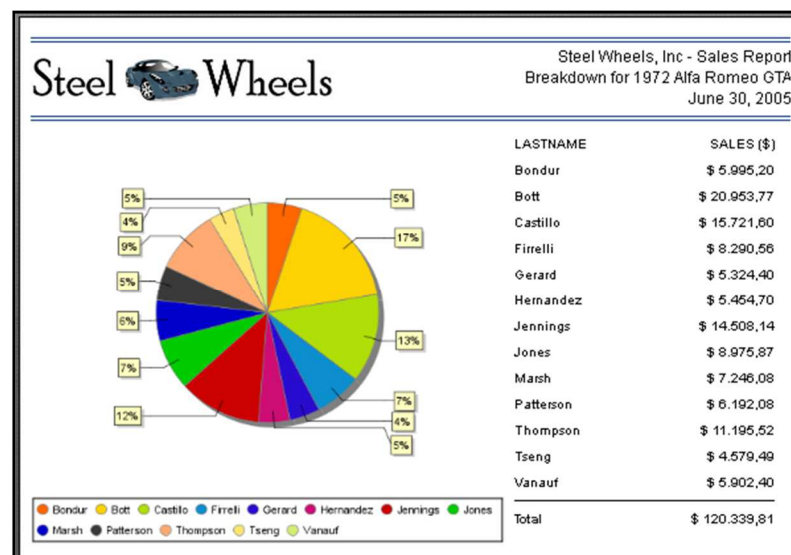


Figura 07 - Relatório de BI
Fonte: o autor.

Os gráficos também são estáticos e podem ter filtros incluídos. Podem ser apresentados vários gráficos processados em um mesmo momento, conforme apresentado na Figura 08.

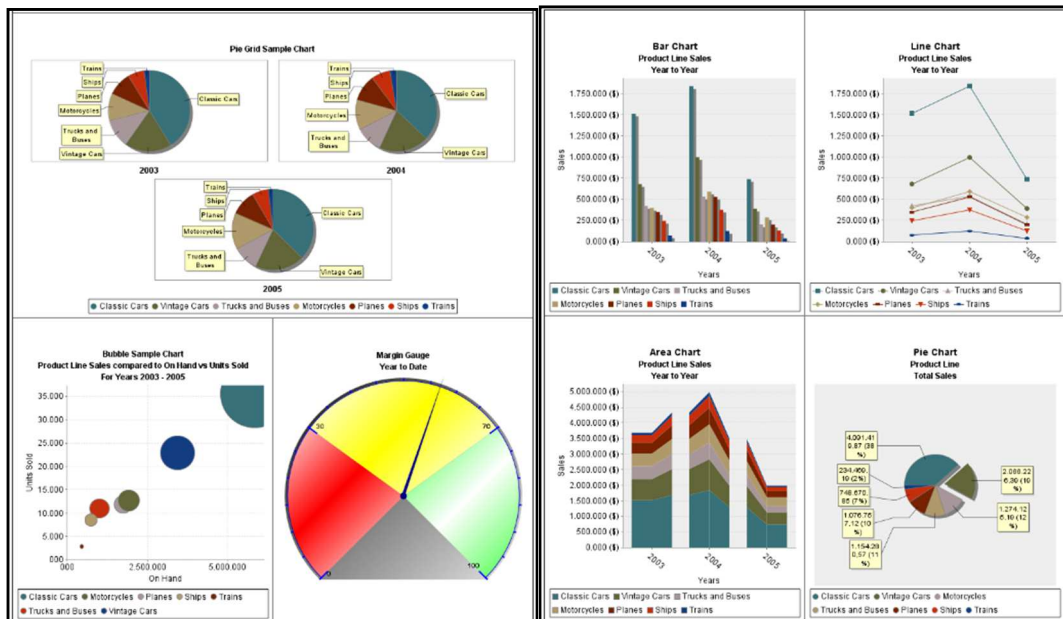


Figura 08 - Relatório de BI
Fonte: o autor.

Os cubos são dinâmicos e podem ter filtros incluídos e tem a sua configuração em tela ajustado durante a apresentação do cubo. É possível fechar e abrir colunas e linhas e inverter ordens de exibição, conforme apresentado na Figura 09.

Arquivo Visualizar Ferramentas Ajuda

Browse

- BI Developer Examples
 - Analysis
 - Business Rules
 - Chart Examples
 - Dashboard Examples
 - Data Integration with Kettle
 - Data Source
 - Printing
 - Reporting
 - Charting in Reports
 - Dependent Parameter E
 - Dynamic Reporting
 - Legacy Steel Wheels Rep
 - PRPT Samples
 - SVS
 - Web Service
 - Steel Wheels
 - Analysis
 - Dashboards
 - Reporting
 - plugin-samples

Google Maps Top 10 Customers Chart Area Chart Chart Examples Revenue Analysis Product Sales Pro

Region Department Positions Measures

Actual	Budget	Variance	Variance Percent	
All Regions All Departments All Positions				
143.639.982,00	143.199.389,00	-440.593,00	-0,31%	
All Regions All Departments All Positions				
6.299.022,00	6.494.166,00	-195.144,00	-3,00%	
Executive Management				
CEO	2.049.625,00	1.988.500,00	-61.125,00	-3,07%
SVP Partnerships	1.962.755,00	1.949.637,00	-13.118,00	-0,67%
SVP Strategic Development	1.061.242,00	1.082.590,00	-21.348,00	-2,00%
SVP WW Operations	1.225.400,00	1.473.459,00	-248.059,00	-16,86%
Finance				
All Positions	12.224.220,00	12.087.406,00	-136.814,00	-1,13%
Administrative Assistant	3.583.061,00	3.467.094,00	-125.967,00	-3,54%
CFO	3.265.672,00	3.167.202,00	-98.470,00	-3,01%
Controller	2.163.973,00	2.160.942,00	-3.031,00	-0,14%
IS	2.164.099,00	2.162.883,00	-2.216,00	-0,10%
Payroll	1.046.415,00	1.139.285,00	-92.870,00	-8,95%
Human Resource				
All Positions	13.075.463,00	12.989.341,00	-86.122,00	-0,66%
Administration	2.049.625,00	2.046.500,00	-3.125,00	-0,15%
Corporate Compliance	1.944.073,00	1.950.032,00	6.019,00	0,31%
EOE	1.942.607,00	1.948.753,00	6.146,00	0,32%
HR Generalists	3.739.190,00	3.589.699,00	-149.491,00	-4,00%
HR Training	1.211.073,00	1.280.532,00	-69.519,00	-5,43%
SVP HR	2.100.955,00	2.170.705,00	-69.750,00	-3,24%
Marketing & Communication				
All Positions	13.910.753,00	13.770.267,00	-140.486,00	-1,02%
Analyst Relations	1.133.375,00	1.212.625,00	-79.250,00	-6,95%
CMO	3.583.061,00	3.460.094,00	-122.967,00	-3,45%
Graphics	3.332.375,00	3.221.125,00	-111.250,00	-3,35%
Press Relations	1.761.226,00	1.786.084,00	24.858,00	1,39%
Product Marketing Mgr	2.942.731,00	2.773.899,00	-168.832,00	-5,74%
Writer	1.257.985,00	1.326.440,00	-68.455,00	-5,16%
Product Development				
All Positions	10.644.102,00	10.786.611,00	-142.509,00	-1,32%
CTO	1.257.985,00	1.326.440,00	-68.455,00	-5,16%
Engineer	2.855.925,00	2.785.950,00	-69.975,00	-2,51%
QA Engineer	1.046.415,00	1.132.285,00	-86.870,00	-7,97%
QA Manager	1.086.463,00	1.163.777,00	-83.314,00	-7,17%
Senior Engineer	3.265.672,00	3.160.202,00	-105.470,00	-3,24%
VP Engineering	1.132.642,00	1.211.957,00	-79.315,00	-6,54%
Professional Services				
All Positions	76.317.649,00	76.098.206,00	-219.443,00	-0,29%
SVP Services	15.421.220,00	15.716.000,00	-294.780,00	-1,88%
Senior Consultant	15.160.199,00	15.316.000,00	-155.801,00	-1,02%
Services Mgr	17.401.786,00	17.916.000,00	-514.214,00	-2,87%
Staff Consultant	14.220.896,00	14.634.206,00	-413.310,00	-2,82%
Trainer	14.112.548,00	12.516.000,00	1.597.548,00	12,76%
Sales				
All Positions	11.168.773,00	10.973.332,00	-195.381,00	-1,73%
Account Executive	1.279.975,00	1.323.525,00	-43.550,00	-3,29%
District Manager	2.782.625,00	2.696.000,00	-86.625,00	-3,11%
Pre-Sales	2.599.375,00	2.528.625,00	-70.750,00	-2,80%
Sales Rep	2.746.975,00	2.662.525,00	-84.450,00	-3,08%
Senior Sales Rep	1.760.823,00	1.762.717,00	1.894,00	0,11%

Arquivos

Quadrant Slice and Dice

Figura 09 – Relatório de BI
Fonte: o autor.

Os *dashboards* são dinâmicos, podem ser ajustados conforme o clique em determinada informação, as demais apresentações se ajustam a estas informações.

6. Big Data



Disponível em: <https://pixabay.com/photo-1667212/>

Estamos vivendo na era das informações volumosas. O volume de informações é medido em exabytes, em uma escala que vai do: bit, byte, kylobyte, megaybte, gigabyte, terabyte, petabyte, exabyte, zettabyte ao yottabyte.

Chamam isto de Big Data (TOLE, 2013), mas anos atrás Korth e Silberschatz (1997) mencionavam sobre isto e chamavam esta nova revolução de "explosão de informações". A comparação dos novos acontecimentos é com as revoluções como a invenção da imprensa por Gutenberg, com a logística das informações a todos os cantos do mundo e ainda a invenção do telefone por Graham Bell, que distribuía a informação imediatamente, em tempo real.

O volume tem aumentado a cada ano e as razões são as diversas:

- O armazenamento de dados hoje é barato, os meios de armazenamento: discos rígidos, DVDs, pendrives e HDs externos estão muito mais baratos e suas capacidades são muito maiores. Também os meios gratuitos, com serviços de hospedagem gratuitos nas nuvens.
- Os aparelhos, os aplicativos e as pessoas estão mais integrados à tecnologia e, conseqüentemente, geram e armazenam mais informações. Crianças de 2 anos ou menores já operam celulares, tablets e outros dispositivos e a terceira idade estão assimilando melhor a tecnologia.
- O hábito de manter informações, e-mails e determinados documentos vai acumulando-os em volumosos arquivos.

- O aumento das possibilidades de serviços para captar, publicar e difundir as informações, assim como os blogs, twitter, e-mail, redes globais, conexões sem fio etc.

Além do grande volume de dados gerados, coletados, armazenados etc., a velocidade de transmissão seja qual for o meio, banda larga por cabo, via fibra ótica, 3G, 4G ou wifi e a diversidade de tipos de informações, assim como planilhas, textos, imagens, sons, entre outros, apoiam o sobrecarregar do ser humano e das organizações.

Além disto, o ser humano passou para a Tecnologia da Informação a sua complexidade. Podemos armazenar dados não estruturados como imagens, vídeos, sons e textos. Tudo isto possibilita análises mais complexas com o desenvolvimento de softwares com funções de Inteligência Artificial. As análises e necessidades evoluíram, com isto o processamento também aumenta. Antes, a necessidade era de encontrar um endereço de um cliente em um Banco de Dados, hoje é necessário saber a faixa de idade que mais consome determinados produtos, em uma determinada faixa de preço, apresentado por loja, cidade e país.

Alguém vai dizer que o volume de informações é bom, porque as pessoas e organizações possuem mais informação para tomar decisões. Por outro lado, com o grande volume geram muitas opções de escolha, com mais informações para ler, mais conhecimento para absorver criando uma dificuldade de trabalhar com tanta informação disponível e ainda de encontrar as informações necessárias.

O BI passa a ter uma função primordial neste volume de informações, permitindo a extração, utilização e auxiliando imensamente na tomada de decisões pelas organizações.

Por dentro dos fatos

Extração, Transformação e Carga de Dados

Existem vários obstáculos que podem ser enfrentados para montagem e execução da Extração Transformação e Leitura (ETL). Os diversos problemas na coleta das informações nas bases transacionais das organizações, além das atividades da área de Tecnologia da Informação (TI), tornam o trabalho mais árduo, pois comumente não é a atividade principal da companhia. Os baixos investimentos que a equipe de TI enfrenta, além dos problemas de toda a companhia que são direcionados para a TI.

A construção de um Data Warehouse (DW) necessita de informações de diversos Sistemas ou outras fontes de Dados. Para compor a nova base, é necessário ter o domínio sobre a linguagem Structured Query Language (SQL) e conhecimento técnico sobre Banco de Dados, isso será o diferencial para montar o processo de ETL. Outro detalhe importante é que o carregamento de informações em um DW deve ter preferencialmente informações resumidas.

Um outro problema é o conceito sobre um determinado termo em algumas áreas da empresa que podem ser confundidos ou mal interpretados.

Para a execução do ETL, estes problemas devem ser superados e as questões definidas antes da execução.

Outro problema é a volatilidade dos dados, pois a base de dados deve ser consistente no momento da extração. Isto significa interromper todas as atualizações que possam influenciar na validação e consistência dos dados, um Database administrator (DBA) pode criar uma cópia da base (*dump*) para garantir esta consistência. Quando não for possível fazer o *dump*, é possível executar um script no início e no final do processo para quantificar e validar se houve alteração durante a carga. A divergência aceitável fica entre 1% e 3% da base original, para que a equipe de negócio valide o script de carga.

Uma dica importante é sempre mapear os registros que não se enquadram em nenhuma regra de negócio, e uma forma de identificar problemas é verificar se estes dados estão acima de 8% do total de dados carregados, se isto ocorrer será importante revisar possíveis erros nos scripts.

No mapeamento das bases, após as especificações das regras é necessário se preocupar com o tempo decorrido para a execução dos scripts sem prejudicar as demais áreas. A carga dos dados deve ser processada por regra em períodos de menor volume de transações nas bases, isto, geralmente, ocorre nos finais de semana, feriados ou nas madrugadas. Por isso, a rotina de testes deve ser exaustivamente aplicada para verificar o volume de dados que será carregado e avaliar o espaço disponível na base para não atrapalhar no andamento da carga.

Antes de executar o script faça uma amostragem da base para avaliação dos que requisitaram o DW, com isto será certificada a informação solicitada.

Outra dica importante, é realizar a criação do ETL em etapas, com os resultados armazenados em tabelas distintas, esta dica agiliza algum reprocessamento que seja necessário, sem prejudicar a janela de processamento em execução, contudo é necessário reservar mais espaço.

Uma ferramenta de ETL agiliza imensamente o processo de carga para DW. Lembrando que este processo pode ser utilizado ainda para agilizar qualquer tipo de conversão, exportação e importação de dados necessários. É necessário despendar algum tempo no desenvolvimento dos scripts, contudo os resultados serão observados em agilidade, acessibilidade bem como a praticidade e manutenibilidade dos processos criados.

Fonte: AprendendoETL ([2016.], on-line).

7. Considerações finais

Nesta unidade, a proposta foi apresentar algumas das principais características que permeiam o uso de sistemas de informação para o gerenciamento de

organizações empresariais. O conceito de BI estabelece uma visão planejada e organizada no uso dos dados e informações presentes na maioria das empresas. De fato, devemos considerar que diante do atual cenário no qual as empresas estão inseridas e também do grande volume de informações que trafegam entre as diversas áreas existentes em cada organização, as ações envolvendo o gerenciamento dos dados e sua disponibilidade são fundamentais para garantir o sucesso da organização.

Com o ambiente atual ampliando a competitividade nos negócios, na qualidade, disponibilidade e na rapidez de acesso às informações, a divulgação e a manutenção das informações estão diretamente ligadas ao desenvolvimento da organização, no sentido de potencializar suas atividades e resultados.

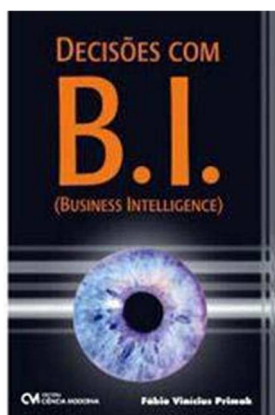
Vimos que a utilização de BI com o uso de múltiplas informações tem proporcionado maior valor e também a melhora no uso de ativos de informações valiosas.

Essas informações abrem caminhos para os profissionais de Bancos de Dados e também de especialistas em BI, trazendo muitas oportunidades aos que se especializarem nestas áreas.

Um DW parte de um BI, também apoia nas tomadas de decisões gerenciais. Os elementos que compõem a visão das condições de uma organização, em um determinado período de tempo e a ideia é que esta estrutura tenha uma infraestrutura de Banco de Dados, on-line e com registro de dados.

O Big Data também é uma questão muito atual e terá o seu fluxo de geração, armazenamento e com alta capacidade de dados e processamento dentro de uma normalidade que será estabelecida com métodos, processos e dispositivos.

Elementos complementares



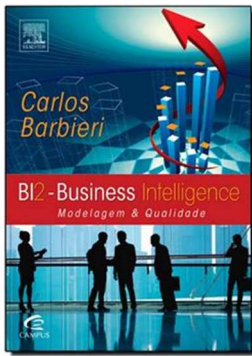
Título: Decisões com B.I. - Business Intelligence

Autor: Fábio Vinícius Primak

Editora: Ciência Moderna

Sinopse: o enfoque principal desta obra é Business Intelligence baseando-se em uma linguagem cotidiana, simples, detalhada e, sobretudo, acessível! É especialmente imprescindível para profissionais das áreas de gestão empresarial, responsáveis pela informática e informações na empresa além de ser

material auxiliar dos mais variados cursos sobre novas tecnologias aplicadas ao desenvolvimento da empresa tendo como base explicitamente a informação correta.



Título: BI2 - Business Inteligente - Modelagem & Tecnologia

Autor: Carlos Barbieri

Editora: Campus

Sinopse: voltado para gestores de tecnologia da informação em grandes corporações, como CIOs, gerentes de BI, analistas e projetistas de sistemas de inteligência de negócios, que planejam organizar e transformar seus dados armazenados em sistemas de informação gerenciais. Esse livro é dedicado aos aspectos de governança e qualidade de dados, como elementos fundamentais para a implementação de projetos de Business Intelligence (BI) bem-sucedidos. Além disso, discute também os novos caminhos do BI, que se amplificará em uma série de novos domínios, como BehaviorIntelligence (BHI), BI aplicado a dados não estruturados, BI para gerência de projetos, BI-Ágil etc. O livro também procura trazer reflexões sobre os novos momentos da era do zettabytes ou big data.

Referências

Aprendendo ETL, Disponível em: < <http://www.aprendendoetl.com.br>>. Acesso em: 04 jun. 2017.

BERTIN, J. **Semiology of Graphics**: Diagrams, Networks, Maps. University of Wisconsin Press, 1983.

TOLE, A. A. **Big Data Challenges**. Database Systems Journal, v. IV, n. 3, 2013, p.31-40.

KORTH, H.; SILBERSCHATZ, A. **Database Research Faces the Information Explosion**. Communications of the ACM, v. 40, n.2, Fevereiro de 1997, p.139-142.

TURBAN, E. **Business intelligence**: um enfoque gerencial para a inteligência do negócio. São Paulo: Bookman, 2009.

TZU, S. **A arte da Guerra**. Adaptação de James Clavell – 38. ed. São Paulo/Rio de Janeiro - Editora Record, 2002.

UNIDADE III - MODELO DE PREPARAÇÃO DE UM DATA WAREHOUSE

Introdução

O que diferencia as gestões são as ferramentas utilizadas por elas. O mercado vem exigindo que a cada dia as informações estejam disponíveis de forma mais rápida e objetiva na mão dos tomadores de decisão para uma possível vantagem competitiva.

Para se obter informações rapidamente e de fácil acesso para assim acelerar o processo de tomada de decisão, surgiu o conceito de Data Warehouse. O Data Warehouse (DW) é a base de dados principal do Business Intelligence (BI). É possível afirmar que BI obtém os dados de instituições e gera informações relevantes à tomada de decisão. É formado por um conjunto de tecnologias que auxiliam a disponibilizar essas informações, tais como o Data Mining (Mineração de Dados), Data Warehouse (Armazém de Dados), e outras, e todas visam gerar conhecimento de acordo com a necessidade.

O DW é composto pelos dados de sistemas transacionais distintos de uma organização, projetado para realizar consultas analíticas e estruturadas para dar suporte à tomada de decisão gerencial. Existe todo um processo para que os dados dos sistemas transacionais estejam disponíveis no DW e também diversas ferramentas que realizam essa etapa. No mercado encontramos um grande número de ferramentas disponíveis para a criação dessa base de dados, porém o custo elevado limita o uso dessas às empresas de grande porte. Por outro lado, a grande ascensão dessa tecnologia fez com que surgissem ferramentas gratuitas ou com o custo baixo e a consequência disso é que as empresas de médio e pequeno porte também podem utilizar o DW.

O fato de ser uma área de grande ascensão e de existirem poucos profissionais qualificados no mercado motiva a buscar conhecimentos em Business Intelligence (BI) e Data Warehouse (DW). Com isto, será apresentado um modelo de conhecimento para a construção de um Data Warehouse para uma empresa.

A partir desse contexto fique atento aos conceitos e à metodologia de um Data Warehouse e ótimo estudo!

1. Cultura e vantagem de um Data Warehouse



Disponível em: <https://pixabay.com/photo-696966/>

Existe uma grande carência de informações confiáveis destacando a necessidade da implementação de um *Data Warehouse*(DW) para identificar tendências futuras, isto torna a empresa mais apta e competitiva.

A grande competitividade do mercado exige que os gerentes obtenham mais informações certas na hora certa, e que estes sejam capazes de prever acontecimentos e preparar-se para tais.

Muitas empresas colhem bons frutos possibilitados pelas soluções de BI, dentre elas podemos citar como casos de sucesso a General Motors do Brasil (GM), o SBT, a Redecard, a ABC Inco, a GVT, dentre outros. (PRIMAK,2008).

No caso da GM, a implementação da solução de BI proporcionou benefícios às áreas de Marketing e Vendas com informações desde um pedido até sua entrega ao consumidor, também às áreas de Manufatura, Finanças e Compras, responsáveis pela compra de materiais, previsão de venda de veículos, análises de vendas on-line, dentre outros. Hoje a empresa comemora o fato de ser mais e rápido e fácil obter o cruzamento das informações do seu dia-a-dia sobre seus negócios. Outro benefício constatado é a facilidade na criação de relatórios, onde o usuário pode definir os filtros, como por exemplo, filtros por região, tempo ou modelos dos veículos comercializados. (PRIMAK,2008)

Essa rapidez e facilidade permitiram aos executivos da empresa receber relatórios diários com as informações do dia anterior, como por exemplo, quanto foi vendido na empresa. (PRIMAK, 2008)

O SBT estava acostumado a trabalhar com planilhas eletrônicas para gerar relatórios, tais como o balanço do mês. Os funcionários passavam dias inteiros inserindo dados nas planilhas e faziam de duas a três horas extras para conseguir entregar os relatórios dentro do prazo. (PRIMAK, 2008).

Quando foi informado a esses funcionários que os relatórios seriam feitos on-line, houve certa resistência por parte destes que ficaram receosos quanto a perder sua função. Para integrar o conceito de BI, foram criados treinamentos e

palestras divulgando as vantagens da utilização desta iniciativa explicando seu papel mais estratégico para a companhia. Até que a resistência foi quebrada e hoje o BI é utilizado por 80 gestores da empresa. (PRIMAK,2008)

A empresa GVT conseguiu reduzir seus custos em até R\$ 3 milhões em 2002, com investimento de R\$ 500 mil. Essa redução se deu apenas com a simplificação e otimização do processo de emissão de relatórios e declaração de tráfego (de usuários) e interconexão (com outras operadoras). Atualmente as áreas que mais utilizam o BI são o departamento de vendas, atendimento ao cliente, engenharia e operações, marketing e financeiro. (PRIMAK, 2008)

Dessa forma, é possível observar as soluções de BI crescendo e apresentando seus benefícios. No mercado atualmente o número de profissionais específicos dessa área é baixo e a demanda pela informação é alta, o que possibilita a concluir é que estudar tecnologias emergentes é uma boa estratégia para se posicionar bem no mercado e se tornar uma referência,principalmente quando se trata de uma tecnologia que traz um grande benefício às empresas e auxilia seus grandes executivos e tomadores de decisões.

2. Data Warehouse



Disponível em: <https://pixabay.com/photo-457334/>

2.1. Definição de Data Warehouse

Entende-se Data Warehouse (DW) como um armazém de dados ou depósito de dados. É uma base de dados que armazena informações corporativas de diferentes fontes de dados, é estruturado para realizar consultas analíticas e projetada especificamente para dar suporte à tomada de decisão gerencial. Os dados de um DW são sumarizados, históricos e detalhados para suportar a análise de negócios e as decisões estratégicas(COLAÇO, 2004).

O modelo de dados utilizado no DW é o dimensional que favorece as consultas de relatórios e análise de grandes volumes de dados.

Existem diversas definições a respeito deste assunto, segundo Inmon(1997), idealizador do conceito, um Data Warehouse “é uma coleção de dados

integrados, orientados por assunto, variáveis com o tempo e não voláteis, usados para dar suporte ao processo gerencial de tomada de decisão”.

Podemos afirmar que o DW é, na verdade, uma coleção de dados derivados dos dados operacionais para sistemas de suporte à decisão (PRIMAK, 2008).

Os bancos de dados operacionais, ou seja, os bancos de dados dos sistemas transacionais da empresa são utilizados por toda a organização para inserir, alterar, consultar ou apagar informações, por isso seus dados podem sofrer constantes mudanças. Como não há ocorrência de redundância nos dados e as informações históricas não ficam armazenadas por muito tempo, este tipo de banco de dados não exige uma capacidade grande de armazenamento (PRIMAK, 2008).

O DW ao contrário dos bancos de dados operacionais, armazenam dados analíticos, destinados às necessidades das gerências no processo de tomada de decisões, podendo envolver complexas consultas que necessitam acessar um grande número de registros. Dessa forma, é importante a existência de muitos índices criados para acessar essas informações históricas de longas datas, e para tal deve ter grande capacidade de armazenamento e processamento (PRIMAK, 2008).

Para o melhor entendimento do conceito de Data Warehouse a tabela 3 ilustra a comparação entre um banco de dados operacionais ou transacionais e um DW:

Tabela 3 – Comparação entre BD Transacionais e Data Warehouse
Fonte: PRIMAK (2008)

Características	Banco de dados Transacionais	Data Warehouse
Objetivo	Operações diárias do negócio	Analisar o negócio
Uso	Operacional	Informativo
Tipo de processamento	OLTP	OLAP
Unidade de trabalho	Inclusão, alteração, exclusão	Carga e consulta
Número de usuários	Milhares	Centenas
Tipo de usuário	Operadores	Comunidade gerencial
Interação do usuário	Somente pré-definida	Pré-definida e ad-hoc
Volume	Megabytes – Gigabytes	Gigabytes – Terabytes
Histórico	60 a 90 dias	5 a 10 anos
Granularidade	Detalhados	Detalhados e Resumidos
Redundância	Não ocorre	Ocorre
Estrutura	Estática	Variável
Manutenção	Mínima	Constante

desejada		
Acesso a registros	Dezenas	Milhares
Atualização	Contínua (tempo real)	Periódica (em batch)
Integridade	Transação	A cada atualização
Número de índices	Poucos/simples	Muitos/Complexos
Intenção dos índices	Localizar registro	Aperfeiçoar consultas

2.2. Características de um Data Warehouse

As características fundamentais de um DW são: Orientação por assunto, Integração, Variação no tempo, Não volatilidade, Localização, Credibilidade dos dados, Granularidade e Metadados. A seguir a descrição detalhada de cada característica.

2.3. DW Orientado por assunto

Ser orientado por assunto significa organizar em temas ou assuntos específicos da empresa de acordo com o interesse e necessidades, contendo apenas informações relevantes ao processo de tomada de decisão. Essa característica é muito importante, pois determina toda a modelagem, que será baseada nesse assunto.

Como exemplo pode-se imaginar uma empresa que trabalha com vendas de produtos alimentícios no varejo e seu maior interesse é o perfil de seus compradores, nesse caso o DW será voltado para as pessoas que compram seus produtos e não para os produtos que são vendidos (JUNIOR, 2004).

2.4. DW Integrado

A integração é uma característica muito importante do DW, talvez a mais importante. Integração significa a padronizar os dados de diversos sistemas, para uma representação única no ambiente DW. Como as fontes de dados do DW são diversos sistemas é normal que todos estejam em formatos diferentes (MACHADO, 2000).

A figura 10 exemplifica como é realizada a integração dos dados.

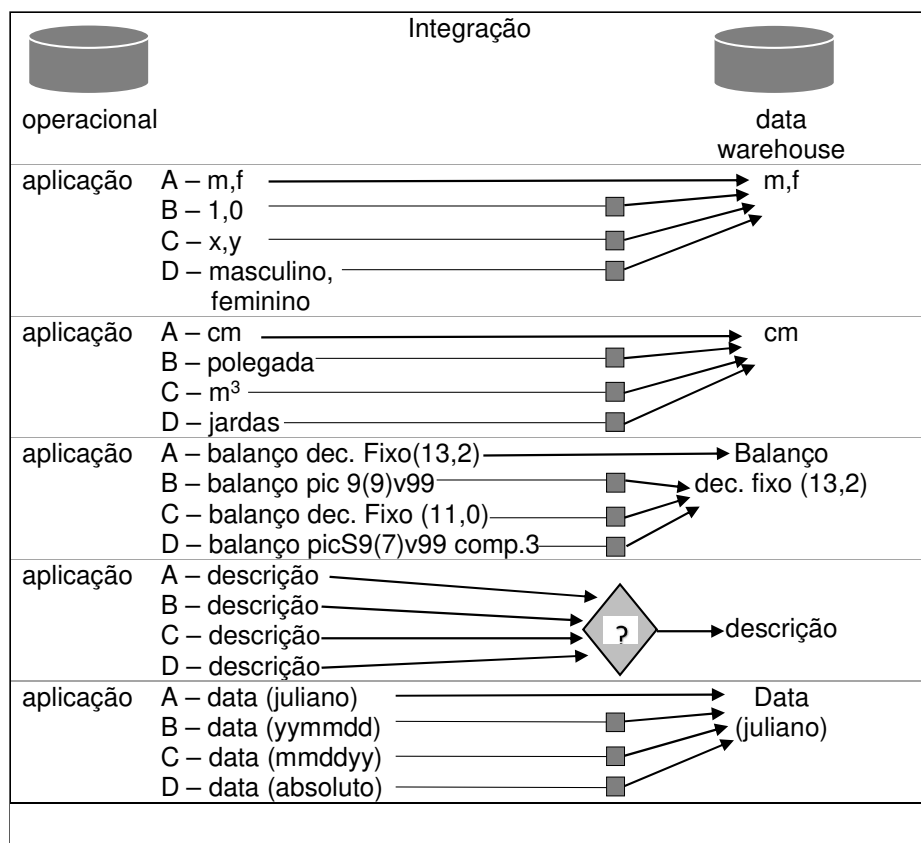
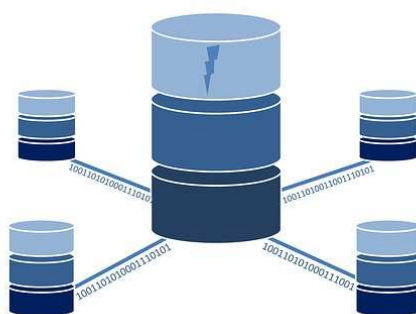


Figura 10 – Exemplo de integração de dados
 Fonte: SOWEK (2010), adaptado pelo autor

2.5. Não-Volátil e Histórico

Nos bancos de dados transacionais, são realizadas quatro operações básicas, tais como inserção, atualização, deleção e consulta. No ambiente de DW os dados após serem integrados, são carregados e armazenados no banco de dados analítico e as operações possíveis para esses dados são apenas de consulta, ou seja, o DW permite apenas a carga inicial dos dados (insert) e consultas a eles (select). Como os usuários não atualizam dados, os bancos de dados analíticos são também chamados de “somente leitura”, o que possibilita a existência de grande volume de dados históricos (MACHADO, 2000).

2.6. Variante no Tempo



Disponível em: <https://pixabay.com/photo-1954920/>

Como visto anteriormente os dados de um DW não são atualizados, somente inseridos, dessa forma os dados representam exatamente aquele determinado instante de tempo, como analogia pode-se utilizar uma fotografia, que representa aquele exato instante, o que proporciona o armazenamento histórico dos dados (MACHADO, 2000).

Segundo Inmon (1997), o DW geralmente armazena informações por um período de cinco a dez anos, já os sistemas operacionais são limitados armazenando os dados entre 60 e 90 dias. Com esse histórico podemos posicionar a empresa estrategicamente para ser mais competitiva, e com isso maximizar os lucros, diminuindo os erros na tomada de decisão eficaz.

2.7. Granularidade

A granularidade refere-se ao nível de detalhe dos dados que serão armazenados num DW, sendo muito importante, pois define o volume de dados que serão armazenados. Por exemplo, se definirmos ter os detalhes por dia, teremos milhares de registros ao longo do mês, mas se caso for definido o detalhe por mês a quantidade de registros diminui absurdamente e conseqüentemente o tamanho do banco de dados (MACHADO, 2000).

Quando se diz que a granularidade de uma tabela é alta, estamos dizendo que o nível de detalhe é baixo, quando o nível de detalhe for alto a granularidade dessa tabela será baixa.

A tabela 3 demonstra a diferença de nível de granularidade dos dados:

Tabela 3 – Diferença no nível de granularidade
Fonte: PRIMAK (2008)

Granularidade Baixa	Granularidade Alta
Maior nível de detalhes	Menor quantidade de detalhes
Grande quantidade de espaço em disco	Economia de espaço em disco
Menor desempenho nas consultas	Melhor desempenho nas consultas
Aumento da capacidade de responder a qualquer questão	Menor capacidade de responder a qualquer questão

A Figura 11 demonstra de uma forma mais intuitiva os níveis de granularidade.

Níveis de Granularidade									
Baixa					Alta				
Dia	Produto	Quantidade	Vir Unitário	Valor Total	Mês/Ano	Produto	Quantidade	Vir Unitário	Valor Total
01/01/2017	A	10	27,00	270,00	jan/17	A	40	27,00	1.080,00
02/01/2017	B	15	45,00	675,00	jan/17	B	93	45,00	4.185,00
02/01/2017	A	30	27,00	810,00					
03/01/2017	B	18	45,00	810,00					
04/01/2017	B	60	45,00	2.700,00					

Figura 11 – Exemplo de Nível de Granularidade
Fonte: o autor

Saiba mais ou Atenção!

A granularidade deve ter um foco de maior atenção, pois é uma das mais importantes definições quando tratamos da modelagem de dados para um Data Warehouse (DW).

A dimensão da informação, seu nível de granularidade deve ser definida no início do projeto e de acordo com as necessidades e envolve a granularidade individual e independente em cada tabela Fato.

Este é um assunto muito importante e que envolve um grande impacto se for mal dimensionado e pode desencadear a inviabilização do projeto.

Saiba mais em: <https://canaltech.com.br/business-intelligence/a-granularidade-de-dados-no-data-warehouse-26310/>, acessado em Julho/2017.

2.8. Metadados

Os metadados são os “dados dos dados” e são peças fundamentais num DW. Além da criação do banco de dados e dos artefatos que o compõem, como dicionário de dados, modelos e etc., criados também no banco de dados tradicional, o Data Warehouse necessita gerar outras documentações com mais informações. (PRIMAK, 2008)

Segundo Colaço (2004), num sistema OLTP os documentos gerados são somente referentes ao levantamento dos dados, banco de dados e sistemas que o alimentam. No DW, além destes são geradas diversas documentações que descrevem melhor sua estrutura, tais como:

- Levantamento de relatórios a serem gerados
- Origem dos dados que alimentam o DW
- Processos de extração, tratamento e rotinas de carga dos dados
- Regras de Negócio da empresa
- Frequência de acesso aos dados

O interesse na área de meta dados está aumentando em função da grande necessidade das organizações conhecerem melhor os dados que a mantêm. De acordo com Colaço (2004), “sem uma documentação eficiente dos dados, é dificultada aos usuários a localização de dados necessários para suas aplicações”.

2.9. Data Mart

Segundo Colaço (2004), um Data Mart ou simplesmente DM é “um subconjunto dos dados contidos em um Data Warehouse extraído para um ambiente separado”. Significa dizer que DM são pequenas porções de dados do DW organizadas por assunto (MACHADO, 2000).

Os dados do DM são direcionados a um departamento ou a uma área específica do negócio. Sua principal vantagem é a possibilidade de retorno rápido, garantindo um maior envolvimento do usuário final, capaz de avaliar os benefícios extraídos de seu investimento (MACHADO, 2000).

2.10. ODS

O conceito de ODS (Operational Data Store) está relacionado ao armazenamento e tratamento de dados operacionais de forma consolidada, assim como nos Data Warehouses e Data Marts, porém sem as características dimensionais. Representam a metade do caminho entre os dados dos sistemas transacionais e o DW. Mesmo não possuindo características dimensionais, os ODS também oferecem informações importantes do ponto de vista decisório, devido a sua característica de consolidação e integração de várias fontes de dados (PRIMAK, 2008).

2.11. OLTP

OLTP (On-line Transaction Processing) ou processamento de transações em tempo real são sistemas que se encarregam de registrar todas as operações de uma determinada organização em tempo real. Seus dados matêm-se usualmente na situação corrente, as atualizações dos dados são realizadas constantemente e seu nível de detalhe é alto (JUNIOR, 2004).

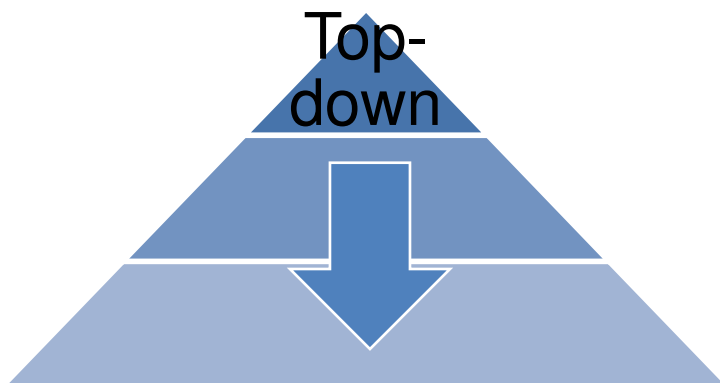
2.12. Tipos de Implementação

Existem várias formas de implementação da arquitetura de DW, as principais são a top down, bottom up e a combinação das duas, chamadas por Machado (2000) de Intermediária.

A escolha de um tipo de abordagem de implementação depende de fatores como a infraestrutura de TI, a arquitetura escolhida, o escopo da implementação, os recursos disponíveis, tanto quanto pessoal capacitado, quanto necessidade de acesso corporativo dos dados, o tempo disponível para implementação e o retorno de investimento desejado (MACHADO, 2000).

A seguir iremos descrever com mais detalhes os três tipos de implementação de uma arquitetura de DW.

2.12.1. Implementação Top Down



Esse tipo de implementação envolve todos os departamentos da empresa e requer um maior planejamento e trabalho de definições conceituais de tecnologia, antes de iniciar o projeto (MACHADO, 2000).

Nessa arquitetura o processo se inicia com a extração, transformação e integração dos dados dos sistemas operacionais para um ODS (Operational Data Store). Após a carregar os dados no ODS, estes são transferidos para o DW. A partir do DW os dados são extraídos para os Data Marts (MACHADO,2000).

Existem vantagens e desvantagens para aplicação dessa arquitetura.

Vantagens:

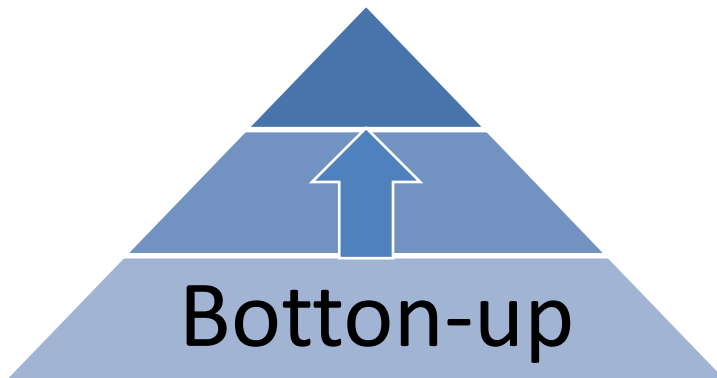
- Garantia da existência de um único conjunto de aplicações para extração, limpeza, e integração dos dados;
- Processos centralizados de manutenção e monitoração;
- Repositório central de metadados, permitindo manutenções mais simples;
- Concentra todos os negócios da empresa, possibilitando extrações em níveis menores de informações.

Desvantagens:

- Implementação longa, pois envolve a interação de todas às áreas da empresa;
- A demora de finalização do projeto e a falta de retorno podem induzir expectativas nos usuários;
- Necessita de uma equipe de desenvolvedores e usuários finais altamente capacitados para avaliar as consultas e informações.

Segundo Machado (2000), essa implementação pode ser bem sucedida se houver uma figura central na empresa da administração de tecnologia da informação. Dessa forma, na cultura de negócios do Brasil, onde as empresas não possuem visão corporativa, e a necessidade de obtenção de resultados é curta, esse tipo de implementação é inviável.

2.12.2.Implementação BottomUp



A popularidade da implementação BottomUp vem crescendo cada vez mais em virtude da dificuldade e custo identificados na implementação TopDown.

Nessa abordagem, é possível criar os Data Marts sem precisar a definição final de um Data Warehouse na empresa. A idéia é a implementação incremental do DW conforme forem sendo realizados os DM. Essa implementação é mais viável e mais escolhida pelos executivos, uma vez que o retorno de investimento é muito rápido.

O processo de implementação bottomup, tem início com a extração, transformação e integração dos dados dos sistemas operacionais para um ou mais Data Marts (MACHADO, 2000).

Segundo Machado (2000), “o propósito desta implementação é a construção de um Data Warehouse incremental a partir do desenvolvimento de Data Marts independentes”.

As desvantagens decorrentes desse método são devido à dificuldade de padronização única de metadados, podendo gerar redundância e inconsistências dos dados. Porém, esse fator pode ser minimizado com um monitoramento, planejamento e principalmente uma metodologia de desenvolvimento (MACHADO, 2000).

Por outro lado as vantagens são bem maiores. A implementação deste tipo de arquitetura é rápida, o retorno também ocorre em curto tempo e isso traz muito mais satisfação ao usuário final (MACHADO, 2000).

2.13. Modelo Dimensional

Modelagem Dimensional é uma técnica voltada especialmente para garantir extrações com alto desempenho, combinando dados históricos em series e permitindo uma visualização dos dados de forma mais clara e objetiva (COLAÇO, 2004).

Segundo Morales (2010) o modelo dimensional é baseado em três elementos:

- Tabelas Fatos
- Tabelas Dimensões

- Medidas

Uma tabela do tipo Fato representa um conjunto de itens composta por dados de medidas e dados de contexto. Cada fato representa um determinado evento do negócio obtido a partir das dimensões (MORALES, 2010).

Uma tabela Dimensão representa contexto em que o fato ocorreu, tais como tempo, produtos, clientes e fornecedores. Ela serve para descrever o fato (MORALES, 2010).

As Medidas são atributos que demonstra o fato na forma quantitativa. O contexto de uma medida, ou seja, a descrição da medida é determinada em função das dimensões que participam do fato (MORALES, 2010).

A visualização dos dados, no modelo dimensional é feita em forma de um cubo, onde cada dimensão do mesmo representa um contexto de um determinado fato e a intersecção entre as dimensões representa as medidas do fato (MORALES, 2010).

A forma simples de organizar os dados no modelo dimensional, além de agilizar o processamento das consultas, permite uma melhor visualização dos dados.

Dentre as vantagens de implementação do modelo dimensional, podemos destacar as seguintes:

- Fácil entendimento para os usuários finais e técnicos;
- Maior agilidade no processamento de consultas;
- Liberdade de mudanças: as tabelas podem sofrer alterações de acordo com a necessidade, como por exemplo adicionando campos novos às dimensões;
- Alterações no modelo não impactam as ferramentas de consulta;

A tabela 4 descreve de uma maneira detalhada e comparativa, os três elementos do Modelo Dimensional.

Tabela 4 - Elementos do Modelo Dimensional
Fonte: MORALES (2010)

FATOS	DIMENSÕES	MEDIDAS
Representam um item, transação ou evento de negócio.	Determinam o contexto de um assunto de negócios, como por exemplo, uma análise de vendas de produtos.	São os atributos numéricos que representam um fato e são determinadas pela combinação das dimensões que participam do mesmo.
Refletem a evolução dos negócios.	São os balizadores de análise de dados.	Representam a performance de um indicador de negócios relativo às dimensões que participam de um fato.
São representados por conjuntos de valores numéricos (medidas) que variam ao longo do tempo.	Normalmente não possuem atributos numéricos, pois são somente descritivas e classificatórias dos elementos que participam de um fato.	Podem possuir uma hierarquia de composição de seu valor.

2.13.1. Tabelas Fatos

É a principal tabela do modelo dimensional, onde estão armazenadas as medidas. O fato representa a medição do negócio, ou seja, tudo aquilo que pode ser medido, como por exemplo, a quantidade de vendas realizadas no mês, ou o número de clientes ativos em determinada empresa. A tabela de dimensões define a granularidade da tabela de fatos, isto é qual é o escopo da medição.

As características básicas de um fato são:

- Variam ao longo do tempo
- Tem valores numéricos
- Seu histórico cresce com o passar do tempo.
- As medições dos fatos devem ser: numéricos e aditivos.

As medidas podem ser classificadas em:

- Valores aditivos: medidas que podem ser aplicadas, operações matemáticas, como soma ou multiplicação, dentre outros.
- Valores não aditivos: medidas que não podem ser manipuladas livremente, como por exemplo, a temperatura ou condição do tempo.

2.13.2. Dimensões

Contém as descrições textuais, referência o contexto em que o fato ocorreu, podemos citar como exemplo de dimensões o tempo, produtos, clientes e fornecedores (MACHADO, 2000).

A maioria dos fatos, conforme figura 12, envolve pelo menos quatro dimensões básicas: onde, quando, quem e o que (MACHADO, 2000).

- Onde: determina o local onde o fato ocorreu (local geográfico, filial).
- Quando: é a própria dimensão tempo.
- Quem: determina que entidades participaram do fato (cliente, fornecedor, funcionário).
- O quê: determina qual é o objeto do fato (produto, serviço).

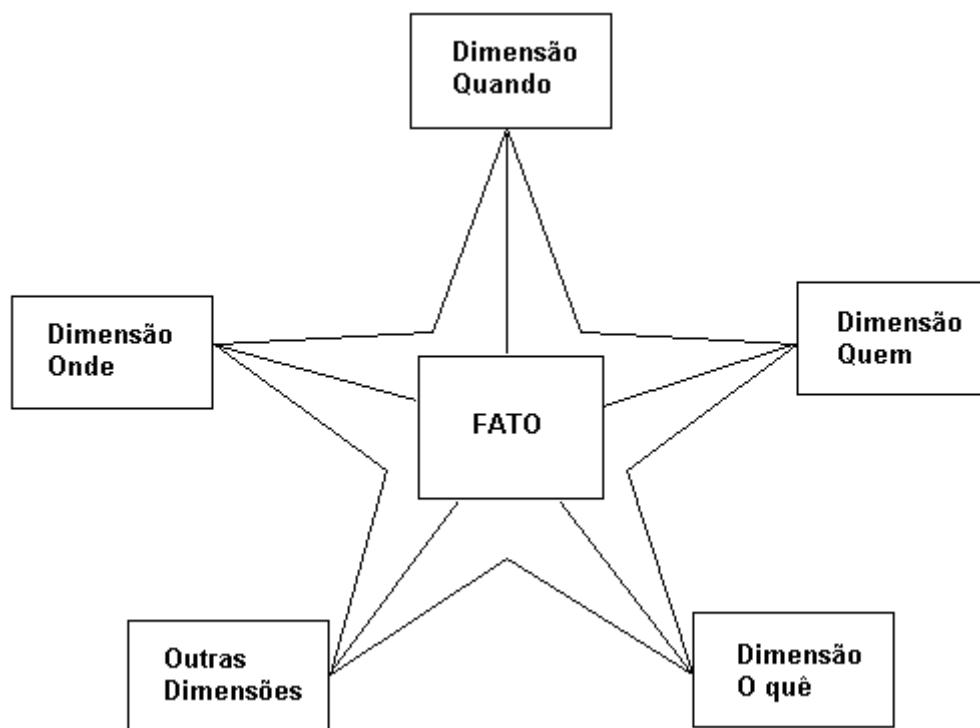


Figura 12 – Dimensões no modelo dimensional
Fonte: MACHADO (2000)

2.13.3. Tipos de Modelagem Dimensional

2.13.3.1. Modelo Estrela (Star Schema)

No modelo estrela todas as tabelas relacionam-se diretamente com a tabela fato, é chamada de estrela porque a tabela fato encontra-se cercada pelas tabelas dimensionais, sendo assim as tabelas dimensionais devem conter todas as descrições necessárias.

Na figura 13 é apresentado a tabela fato VENDA e as dimensões TEMPO, REGIÃO, PRODUTO, VENDEDOR e CLIENTE.

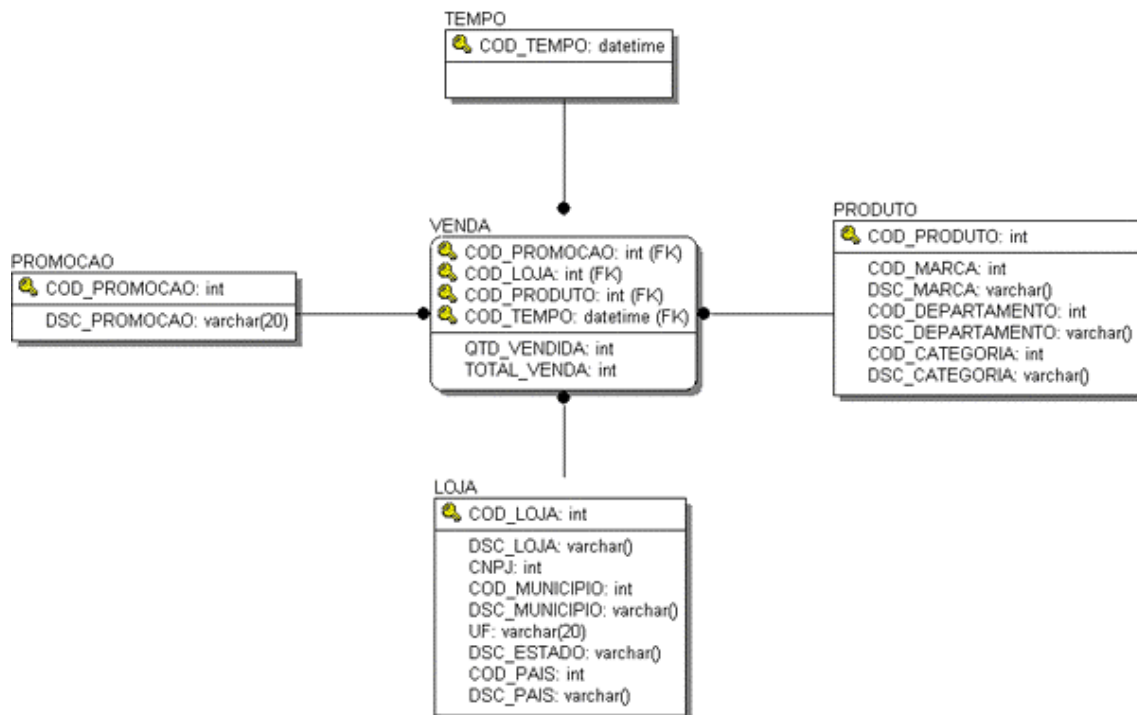


Figura 13 – Representação do modelo estrela
Fonte: MACHADO, 2000

Pontos positivos desse modelo são:

- Simplicidade
- Facilidade de navegação pelos softwares
- Rapidez nas consultas

Pontos Negativos:

- Desperdício de espaço

2.12.3.2. Modelo Floco de Neve (Snowflake)

No modelo Floco de Neve as tabelas dimensionais relacionam-se com a tabela de fatos, mas algumas dimensões relacionam-se apenas entre elas, ou seja, por questão de espaço são criadas tabelas auxiliares que se relacionam com suas tabelas principais.

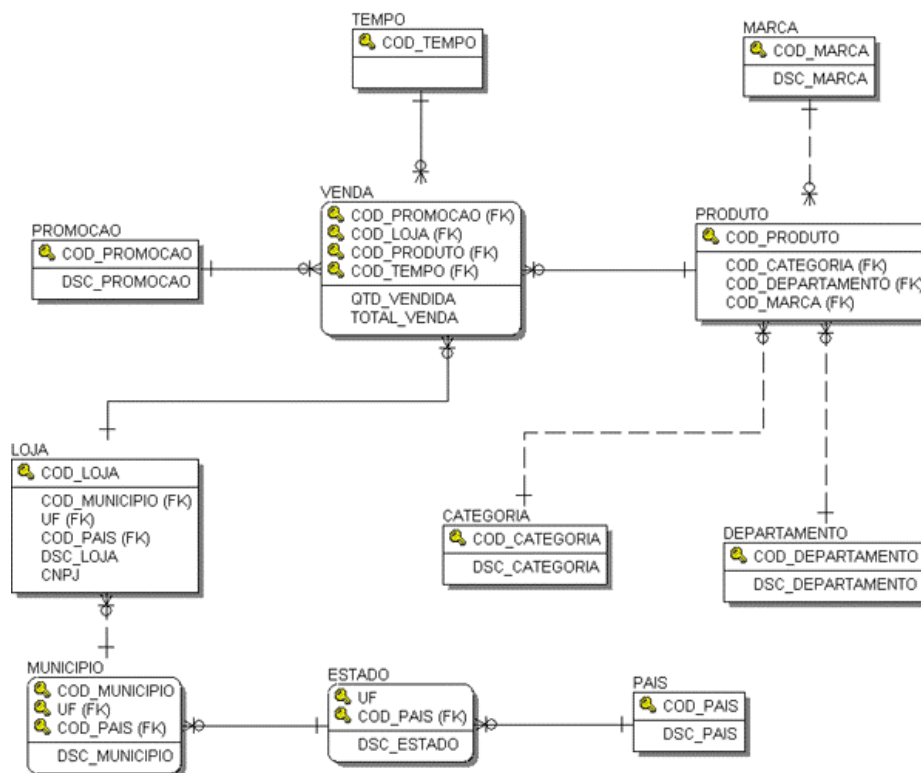


Figura 14 – Representação do modelo Floco de Neve
Fonte: MACHADO, 2000

Nesse modelo existem tabelas de dimensões auxiliares que normalizam as tabelas de dimensões principais.

Assim utilizamos mais tabelas, mas ocupando menos espaço em disco, o modelo fica mais complexo e com a navegação um pouco difícil pelos softwares e conseqüentemente tendo uma resposta um pouco mais lenta quando comparada com o modelo estrela.

2.13. Processo ETL (Extração, Transformação e Carga)

Para que os dados estejam disponíveis em um ambiente de DW é necessário que estes tenham passado por um processo de extração, transformação e carga, conhecido com processo ETL (Extract, Transform e Load). Essas três etapas são de extrema importância para o processo de criação de um DW.

Este é o processo mais importante e poderoso das ferramentas de B.I. pois é capaz de fazer a junção de várias fontes de dados em apenas um repositório e tem que ter seu processo transparente e seguro para tornar o resultado confiável.

A extração é a etapa inicial de processo ETL, onde são selecionados os dados de diversas fontes, bem como arquivos texto, bancos de dados relacionais (mysql, db2, oracle e etc.) e etc. Nesta fase os dados são identificados de acordo com o interesse para o DW.

A etapa de transformação refere-se ao tratamento desses dados, é a segunda etapa realizada durante o processo ETL e a mais crítica. Durante a etapa de transformação, diversos tipos de transformações podem ocorrer:

- Integração,
- Limpeza e Validação,
- Derivação,
- Sumarização,
- Atualizações do Histórico e
- Ordenação.

Resumidamente é a etapa de garantia da integridade dos dados. A última fase do processo ETL é a carga, que consiste em carregar os dados trabalhados no DW.

O objetivo de um processo ETL é fornecer aos usuários do DW dados estruturados, consistentes e de qualidade, de forma a auxiliar na tomada de decisão dos processos de uma organização. O processo ETL é importante devido ao fato de que se um dado não for manipulado corretamente a qualidade da análise cai e conseqüentemente as decisões tomadas podem ser equivocadas e afetar diretamente nos negócios da empresa, trazendo prejuízos a ela.

Para executar essa etapa do DW existem diversas ferramentas no mercado que auxiliam o desenvolvedor. Cada ferramenta tem sua particularidade e sua escolha depende da necessidade de cada empresa. Como exemplos de ferramentas ETL existentes no mercado, existem as seguintes:

- Business Objects Data Integrator;
- DataStage – IBM;
- ETI;
- ODI– Oracle Data Integrator;
- Informática Power Center;
- Pentaho (Open Source);
- Talend;
- TransformOnDemand;
- SQL Server Integration Services (SSIS);
- Sunopsis;
- Syncsort.

Na web

A ETL é um processo muito poderoso, pois realiza o processo de importação, transformação, integração e atualização de informações. Muitas são as ferramentas, e todas integram um alto poder de transformação. O Grupo Gartner faz a pesquisa para identificar as ferramentas mais poderosas.

No site da DEVMEDIA há mais detalhes sobre esta pesquisa e também uma breve apresentação das duas ferramentas mais votadas na pesquisa.

Acesse o link disponível em: <http://www.devmedia.com.br/etl-quais-as-ferramentas-mais-poderosas-do-mercado/6727>, acessado em: julho/2017.

OLAP

OLAP ou On-Line Analytical Processing pode ser traduzido como Processo Analítico On-Line. É a tecnologia que permite ao usuário um rápido acesso para visualizar e analisar os dados com alta flexibilidade e desempenho. Esse alto desempenho se dá graças ao modelo dimensional, que simplifica o processo de pesquisa.

Segundo Machado (2000) é o processo que permite o usuário manipular a visualização das informações gravadas no DW. Essas visualizações são possíveis graças à estrutura desenvolvida para que o usuário obtenha a informação de várias formas possíveis. O tipo de visualização é definido pelo tomador de decisão ou a pessoa responsável pela extração dos dados de acordo com sua necessidade podendo facilmente gerar relatórios diferenciados com informações dinâmicas como gráficos, planilhas entre outros.

Uma técnica de análise multidimensional é possível a navegação na informação de forma a chegar-se do maior ao menor nível de detalhe de forma bem simples e para isso usamos as operações básicas: Drill Down, Drill Up, Slice and Dice.

Drill Down

Operação usada quando existe necessidade de um nível de detalhe maior. Ex: Quando geramos um relatório anual de vendas temos a informação que a empresa conseguiu atingir um dos seus maiores objetivos que é o lucro e para detalhar melhor alteramos o relatório para um dado mensal assim podemos ver com mais detalhes quais os meses tivemos os melhores desempenhos.

Roll Up

Processo contrário do Drill Down necessita-se de uma quantidade menor de detalhes. Ex: Diminuindo o nível de relatório mensal para anual.

Slice and Dice

Possibilita a visualização apenas de uma parte do cubo reduzindo o volume de dados e agilizando as operações de consulta a tabela abaixo descreve as principais diferenças entre os processamentos OLAP e OLTP.

Tabela 5 – Diferenças entre OLAP e OLTP
Fonte: MACHADO (2000)

OLAP	OLTP
Relevância para dados históricos	Mantém usualmente a situação corrente.
Necessidade de ver os dados sob diferentes perspectivas: aplicações dinâmicas	Voltado para velocidade e automação de funções repetitivas
Atualizações quase inexistentes, apenas novas inserções	Atualizações em grande número
Baseado em dados históricos, consolidados e frequentemente totalizados	Baseado em transações
Operações de agregação e cruzamentos	Alto nível de detalhe

Como exemplo de ferramentas OLAP disponíveis no mercado, é possível citar:

- Business Objects;
- Cognos;
- Hyperion;
- Microstrategy;
- MV Business AnalyticsSuite;
- Oracle BI Enterprise Edition;
- Pentaho;

O DW pode ser construído em qualquer base de dados. O BI Server, por exemplo o servidor da Suite Pentaho é multi-plataforma. Ele é o responsável pelo controle de acesso e compartilhamento dos Cubos OLAP, Relatórios e Painéis construídos com as ferramentas de Pentaho com base no DW. O usuário final pode acessar essas informações de diversas formas, como computadores, celulares e etc. (SOUZA, 2010)

A tabela 6 descreve a proposta de metodologia, baseada no processo unificado, que será utilizada neste projeto para o desenvolvimento do Data Warehouse. São descritas em cada fase, as ações ou artefatos de saída que a compõem. O intuito da criação de uma metodologia é possibilitar aos profissionais envolvidos na tarefa de construção de sistemas de DW, uma referência metodológica para auxiliá-los no trabalho de implementação do produto final.

Tabela 6 – Metodologia de Desenvolvimento de DW

Metodologia de Desenvolvimento de DW		
Fase	Ação / Artefato	Descrição
Concepção	Definir o Plano Geral do Projeto	Descrever de forma geral as funções que o data warehouse irá disponibilizar. Também são definidos de forma preliminar a alocação de recursos (máquinas, ferramentas, pessoas) e os prazos.
	Definir o Escopo do Projeto	Consiste no entendimento dos conceitos e tecnologias relacionadas a necessidade de construção do DW.
	Determinar a abordagem de desenvolvimento	Consiste em determinar qual estratégia de desenvolvimento de DW será utilizada. A escolha entre bottom up e top down, determina o tempo e custo de implementação.
Elaboração	Levantamento de Requisitos	Identifica de forma detalhada, com funciona o processo atual, e quais as necessidades identificadas.
	Levantamento dos Relatórios	Identifica a camada de visualização para o usuário final na ferramenta OLAP.
Construção	Modelo de Dados Conceitual	Consiste em uma visão conceitual das informações que serão disponibilizadas no DW, de acordo com a necessidade levantadas.
	Modelo de Dados Lógico	O modelo lógico define como o banco de dados será implementado em um SGBD específico.
	Modelo de Dados Físico	Construção dos scripts sql para implementação do Modelo Dimensional.
	Criação de Metadados	Documento que identifica a origem dos dados que são carregados no Data Warehouse. Também possuem informações de dicionário de dados.
	Construção do Processo ETL	No caso de soluções ETL de mercado são disponibilizados dentro da própria ferramenta ou podem ser objetos de desenvolvimento (Código fonte, shell script, Scripts SQL, etc.).
	Construção dos Relatórios OLAP	Camada de visualização para Usuário final na ferramenta OLAP, artefato disponível dentro da própria solução. Relatórios solicitados pelo requisitante podem ser disponibilizados dentro da própria solução ou exportados em diferentes formatos.
Transição	Manual do Usuário	Manual de utilização da solução para o usuário final. Descreve os procedimentos para abrir os relatórios na ferramenta OLAP.

3. Considerações finais

O Data Warehouse é um banco de dados multidimensional, voltado para sistemas OLAP. É uma base de dados central e histórica que pode armazenar informações de todos os departamentos da empresa. Por intermédio dessa base é possível ler e extrair vários dados de diferentes ângulos, por exemplo: verificar quais os clientes mais rentáveis, os mais propensos à aquisição de novos produtos ou serviços e quais geram mais valor ao longo do tempo.

E para que a tomada de decisões ocorra através dos sistemas OLTP (On-Line Transaction Processing) ocorra, é preciso ter a consciência de que não são tão ágeis, pois os objetivos não são de análise e sim de realizar transações em tempo real. É importante que exista um repositório próprio de dados já consolidados e transformados em informação, separado do ambiente transacional. Isso porque quando se quer realizar uma análise o volume de

dados é bem alto e para isso é necessária uma capacidade maior de processamento dos computadores. Juntar as duas operações (transações em tempo real e análises) em uma máquina comprometeria o desempenho dos sistemas.

Por isso, os sistemas OLAP (On-Line e Analytical Processing) são a alternativa para os sistemas transacionais, devido a oferecer uma visão de dados orientada à análise, além de possuir uma navegação rápida e flexível. Estes sistemas são preparados para realizar relatórios complexos de forma simples e a visão dos dados é multidimensional, ou seja, vai além da visão em duas dimensões de linhas e colunas.

As empresas necessitam, além de conhecer seus clientes, processos e resultados, devem desenvolver estratégias prévias, para solucionar questões gerenciais antecipadamente, deforma a prever tendências futuras, analisar antecipadamente a venda, identificar possíveis novos clientes, analisar o perfil de compra destes, avaliadas vendas de cada funcionário, segmentando-os e definindo incentivos para aumentar as vendas realizadas, ou seja, além do relacionamento Cliente x Concessionária, obter uma visão como um todo dos processos da empresa.

Elementos Complementares



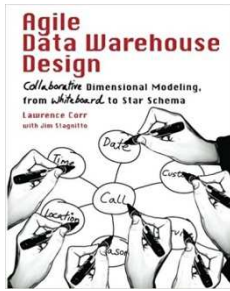
Título: Tecnologia e Projeto de Data Warehouse

Autor: Felipe Nery Rodrigues Machado

Editora: Érica

Sinopse: Com apresentação técnica diferenciada e metodologia embasada na experiência em processos de Data Warehousing, o professor e consultor Felipe Machado transmite seus conhecimentos de forma gradativa e efetiva. Com destaque a aspectos conceituais e orientação à gestão de negócios, as arquiteturas e tecnologias envolvidas no processo de

Data Warehousing são explanadas por meio de exemplos e estudos de caso, considerando-se a versão 2008 do SQL Analisys Server e os conceitos estratégicos de BI Competence Center. Na sexta edição, são apresentadas definições e tecnologias de Big Data e sua correlação com Data Warehouse, para que o leitor sempre acompanhe um padrão de conteúdo atualizado de tecnologias.



Título: Agile Data Warehouse Design: Collaborative Dimensional Modeling, from Whiteboard to Star Schema

Autores: Lawrence Corr e Jim Stagnitto

Editora: Decision one Consulting

Sinopse: Agile Data Warehouse Design é um guia passo a passo para obter dados para requisitos de data warehousing/business intelligence (DW/BI) e direcioná-los para modelos dimensionais de alta performance da melhor maneira.

Referências

COLAÇO, Methanias Jr. **Projetando Sistemas de Apoio a Decisão Baseados em Data Warehouse**. Rio de Janeiro: Axcel Books do Brasil Editora, 2004.

INMON, W. H. **Como construir o Data Warehouse**. Tradução da segunda edição. Rio de Janeiro: Campus, 1997.

MACHADO, Felipe Nery Rodrigues. **Projeto de Data Warehouse: Uma Visão Multidimensional**. São Paulo: Érica, 2000.

MORALES, AranBeyTcholakian. **Sistemas de Apoio a Decisão – Negócios Inteligentes na WEB**. Disponível em <http://www2.stela.ufsc.br/aran/sad/sad_aula4.htm>. Acesso em: Julho de 2017.

PRIMAK, Fábio Vinícius. **Decisões com B.I. (Business Intelligence)**. Rio de Janeiro: Editora Ciência Moderna Ltda., 2008.

SOUZA, Caio Moreno de. **Integração de ferramentas de código aberto (java, pentaho e android) e mapas, aplicada a projetos de inteligência de negócios**. 2010. 57 f. Monografia (Especialização em Tecnologia Java) – Programa de Pós-Graduação em Tecnologia, Universidade Tecnológica Federal do Paraná. Curitiba, 2010.

SOWEK, Carlos Alberto. **O que é Data Warehouse**. Disponível em <<http://www.batebyte.pr.gov.br/>>. Acesso em: Julho de 2017.

UNIDADE IV – O PODER DA FERRAMENTA ETL

Introdução

Extrair, transformar e carregar (ETL) é o processo central de integração de dados e normalmente é associado ao armazenamento de dados. As ferramentas ETL extraem dados de uma fonte escolhida, transformam-na em novos formatos de acordo com as regras de negócios e, em seguida, carregam-na na estrutura de dados de destino.

O objetivo principal é o gerenciamento das regras e processos da crescente diversidade de fontes de dados, os altos volumes de dados processados que o ETL deve executar, o controle do gerenciamento, do desempenho e do custo, são os principais desafios para os usuários.

O ETL é um processo chave para reunir todos os dados em um ambiente padrão e homogêneo. As funções do ETL remodelam os dados relevantes dos sistemas de origem em informações úteis para serem armazenadas no data warehouse.

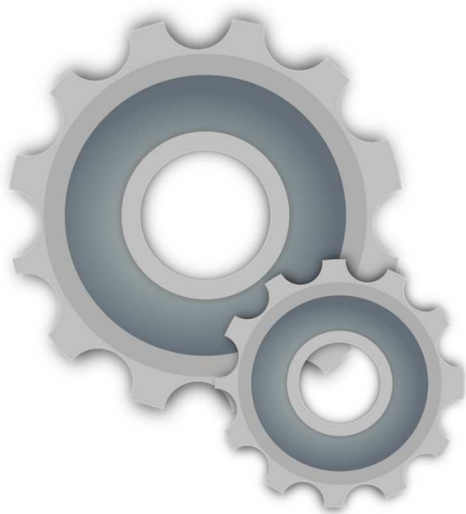
Para isto, também é necessário conhecer as ferramentas disponíveis e realizar uma análise comparativa das várias características das ferramentas de ETL principais. Além disso, a avaliação dessas ferramentas de ETL deve ter como base alguns critérios que determinam ser adequados para uma ferramenta de ETL ter. Através desta comparação, é possível fornecer um conhecimento inicial para avaliar as alternativas líderes de mercado. Outras ferramentas também estão disponíveis, mas foram escolhidos apenas os líderes de mercado, conforme relatório do Grupo Gartner 2017 (BEYER, et al., 2017).

Sem conhecer as ferramentas e suas funções não haveria informações estratégicas no data warehouse. Se os dados de origem extraídos de várias fontes não forem limpos, extraídos corretamente, transformados e integrados da maneira correta, o processo de consulta, que é a espinha dorsal do data warehouse não pode acontecer (SWEIGER et al., 2002).

Também é possível aumentar a velocidade de extração, transformação e carga com suporte de consulta Cache. Como o processo de consulta é a espinha dorsal de um data warehouse, isto permite uma redução de tempo de resposta e melhorará o desempenho esperado.

É importante compreender a abordagem do resultado final esperado, pois as finalidades do ETL podem abordar não somente um data warehouse, mas exportações específicas e com necessidades diversas.

1. A ETL



Disponível em: <https://pixabay.com/photo-145804/>

ETL é uma função de integração de dados que envolve extrair dados de fontes externas (sistemas operacionais), transformá-lo para atender as necessidades de um negócio e, finalmente, carregá-lo em uma outra base de dados unificada ou em um data warehouse.(GONÇALVES, 2012)

Para resolver o problema de unir as várias bases em diversos formatos, as empresas usam a extração, transformação e carregamento (ETL), que inclui ler dados destas fontes, limpá-los, formatá-los uniformemente, e depois gravá-los no repositório de destino a ser explorado. Os dados utilizados nos processos ETL podem vir de qualquer fonte: um aplicativo mainframe, um aplicativo ERP, uma ferramenta CRM, um arquivo data-file ou uma planilha do Excel. (GONÇALVES, 2012)

De acordo com alguns especialistas da indústria, cerca de 60 a 80% do esforço de um projeto de data warehousing é gasto apenas nesse processo. No alto volume de hoje, as técnicas de aquisição de dados do ambiente cliente / servidor têm que coordenar operações de encenação, filtragem, rotinas de limpeza de dados, transformação de dados e técnicas de carga de dados, além de cooperar com a tecnologia de rede para preencher o data warehouse e as bases de dados operacionais (GOUR, 2010).

As funções ETL remodelam os dados relevantes dos sistemas de origem em informações úteis para serem armazenadas no data warehouse. Sem essas funções, não haveria informações estratégicas no data warehouse. Se os dados de origem extraídos de várias fontes não forem limpos, extraídos adequadamente, transformados e integrados de maneira correta, o processo de consulta de um data warehouse não seria possível (SHIM, et al., 1999).

É mais do que apenas um processo de aquisição de dados. Embora a aquisição de dados seja o processo predominante usando as ferramentas ETL, o processo de entrega de dados e o movimento dos dados das funções analíticas ou de sistemas operacionais também usam o processamento ETL. O

conjunto completo de operações ETL deve combinar em um sistema integrado e coeso.

O sistema garante que cada processo se encaixe no esforço geral de forma eficiente, determina como a ferramenta será usada para cada componente e sincronizará todos os eventos do ETL. Deve ter um perito em ETL envolvido no processo que garanta a consistência esperada (GOUR, 2010).

Extrair, Transformar, Carregar são três funções que são combinadas em uma única ferramenta que automatiza o processo para extrair dados de um banco de dados e colocá-lo em outro banco de dados. As funções são descritas a seguir (SATAPATHY, 2006):

- Extrair: o processo de leitura de dados de um ou vários bancos de dados de origem especificado, ou ainda de diversas outras fontes de dados e extrair um subconjunto de dados desejados.
- Transformar: o processo de converter os dados extraídos / adquiridos de sua forma anterior para a forma em que precisa estar para que possa ser colocado em outro banco de dados. A transformação ocorre usando regras ou tabelas de pesquisa ou combinando com outros dados.
- Carregar: o processo de escrever os dados no banco de dados de destino. Uma abordagem alternativa para a integração da informação é a da mediação: os dados são extraídos de fontes de dados originais sob demanda quando uma consulta é posta, com transformação para produzir um resultado de consulta (NUTT et al., 1998).

2. Papel e responsabilidade da ETL



Disponível em: <http://photorack.net/index.php?action=showpic&cat=126&pic=4865>

O arquiteto ETL deve ter os seguintes papéis e responsabilidades (ELMASRI; NAVATHE, 2011).

- O arquiteto ETL deve estar atento às necessidades e requisitos da organização. Este profissional deve entender o ambiente operacional global e os requisitos de desempenho estratégico do sistema proposto. O arquiteto deve interagir com a equipe operacional e técnica do sistema fonte, o administrador do banco de dados do projeto (DBA) e os arquitetos de

infraestrutura técnica para desenvolver o método mais eficiente para extrair dados de origem, identificar o conjunto adequado de índices para as fontes, arquiteto da plataforma de teste, criar bases de dados intermediárias necessárias para uma transformação de dados eficiente e produza a infraestrutura de programação para uma operação ETL bem-sucedida.

- Um programador de ETL não deve ver apenas seu único conjunto de programas, deve ter uma visão mais abrangente. O arquiteto deve ver todo o sistema de programas, ele deve garantir que a equipe técnica entenda o design do banco de dados de destino e seu uso para que as transformações que convertem os dados de origem nas estruturas de dados alvo sejam claramente documentadas e compreendidas. O arquiteto ETL supervisiona todos e cada um dos componentes ETL e seus subcomponentes.

- O processo ETL é muito mais do que o código escrito para mover dados. O arquiteto ETL também serve como ponto central para entender os vários padrões técnicos que precisam ser desenvolvidos se eles ainda não existem. Estes podem incluir limites no tamanho do arquivo ao transmitir dados através da intranet da empresa, requisitos para passar dados através de firewalls que existem entre ambientes internos e externos, padrões de design de dados, padrões de uso de ferramentas de projeto lógico e físico e gerenciamento de configuração de código-fonte, executáveis e documentação. O arquiteto ETL também deve garantir que o processo de design ETL é repetível, documentado e colocado sob o controle adequado de mudanças.

- Uma consideração fundamental para o arquiteto ETL é reconhecer as diferenças significativas que os métodos de projeto e implementação de um sistema de inteligência de negócios têm de uma abordagem de sistema de processamento de transações on-line (OLTP).

- O papel do arquiteto ETL também se estende ao de consultor para o esforço de programação. O arquiteto trabalha em estreita colaboração com os programadores para responder perguntas e desempenhar um papel fundamental na resolução de problemas. Dependendo do tamanho do esforço de programação e da organização do projeto, o arquiteto ETL também pode supervisionar o desenvolvimento das especificações de programação. Em qualquer caso, o arquiteto ETL desempenha um papel fundamental como revisor e aprovador durante o processo de revisão pelos pares.

- Um último papel para o arquiteto ETL deve ser garantir que as várias ferramentas de software necessárias para executar os diferentes tipos de processamento de dados sejam adequadamente selecionadas. O ETL é um dos mais importantes conjuntos de processos para sustento e manutenção da arquitetura e estratégia de Business Intelligence (ELMASRI; NAVATHE, 2011).

3. Problemas de ETL



Disponível em: <https://pixabay.com/photo-989125/>

Existem inúmeros problemas para implementar processos ETL eficientes e confiáveis (ZODE, s.d.). Vejamos alguns deles:

- Desafios técnicos movendo, integrando e transformando dados de ambientes diferentes;
- Janelas de carga curta, tempos de carregamento longos;
- Base de dados inconsistente, difícil de manter as regras de negócios;
- Falta de exposição das regras de negócios aos usuários finais;
- Sistemas de origem que faltam certos dados críticos;
- Pobre desempenho da consulta;

As funções ETL remodelam os dados relevantes dos sistemas de origem em informações úteis para serem armazenadas no data warehouse. Sem essas funções, não haveria informações estratégicas no data warehouse. Segundo Sultan, Aziz (2010) se os dados de origem tirados de várias fontes não forem limpos, extraídos adequadamente, transformados e integrados de maneira adequada, o processo de consulta que é a espinha dorsal do data warehouse não pode acontecer. (SHIM, ET al., 1999)

Na web

Durante o processo de homogeneização, são encontrados muitos conflitos de modelagem. Estes conflitos podem ser divididos em semânticos e estruturais.

Os conflitos semânticos são todos aqueles que envolvem o nome ou a palavra associada às estruturas de modelagem, por exemplo, mesmo nome para diferentes entidades ou diferentes nomes para a mesma entidade. Já os conflitos estruturais englobam os conflitos relativos às estruturas de modelagem escolhidas, tanto no nível de estrutura propriamente dita como no

nível de domínios. Os principais tipos de conflitos estruturais são aqueles de domínio de atributo que se caracterizam pelo uso de diferentes tipos de dados para os mesmos campos.

Para saber sobre os conflitos típicos de domínio de atributo, continue lendo em: http://www.fsma.edu.br/si/Artigos/V2_Artigo1.pdf

4. Categorias de ferramentas de ETL

As ferramentas de ETL podem ser classificadas em duas grandes categorias (ZODE, s.d.):

a) Processo de ETL baseado em codificação manual: Ferramentas de ETL que são desenvolvidas em Perl, COBOL, C e PL / SQL para extrair dados de vários arquivos de origem, transformar os dados e carregar os bancos de dados de destino. Os programas escritos usando este método são longos e difíceis de documentar.

O desenvolvedor ETL tem que usar diferentes linguagens de programação para executar a tarefa ETL, tais como: scripts Perl para extrair os dados a partir de sistemas de origem, realizando transformações, e leitores SQL com procedimentos em PL / SQL utilizados para carregar os dados no data warehouse destino.

As ferramentas de ETL codificadas manualmente tem a vantagem de que os metadados criados podem ser gerenciados diretamente e eles dão a flexibilidade para o desenvolvedor manipular as novas necessidades e os testes de unidade ficam muito mais fáceis. No entanto, suas limitações também existem. Para atender às mudanças contínuas nos grandes volumes de dados gerados através de várias fontes que precisam ser modificados, a frequência provoca uma carga sobre o projeto global.

O ETL codificado manualmente geralmente é lenta execução, enquanto que as modernas ferramentas baseadas executadas em motores de alta velocidade.

b) Ferramenta ETL: Uma vez que, as ferramentas de codificação manual envolvem despesas gerais e são lentas em execução, muitos fornecedores desenvolveram ferramentas a serem adquiridas pelas organizações. Estas ferramentas de ETL começaram a partir de extrações simples em mainframes para direcionar banco de dados e hoje em dia elas estão disponíveis na GUI com funcionalidades adicionais e performances.

As ferramentas de ETL existentes oferecem recursos de transformação, suporte a múltiplos arquivos de entrada ou saída de banco de dados, arquivos, projetos multidimensionais, geração de chave, várias funções de transformação e da base de dados nativa.

As ferramentas têm repositórios de metadados internos que podem diferenciar do repositório de metadados de um data warehouse. Eliminam a sobrecarga de desenvolvimento e manutenção de rotinas complexas de transformações nos fluxos de trabalho de ETL. Além disso, as ferramentas fornecem interface gráfica amigável que permite que o desenvolvedor trabalhe de forma mais tranquila. Estas ferramentas também têm características tais como: monitoramento, agendamento, carregamento em massa, agregação incremental, etc.

As ferramentas ETL de hoje podem ainda ser classificadas em quatro subcategorias (MADSEN, 2004):

i. Ferramentas puramente ETL: são produtos independentes do banco de dados e da ferramenta Business Intelligence com a qual será usado. As empresas não dependem de nenhum outro produto para as funcionalidades oferecidas por elas e também permitem a migração para diferentes bancos de dados sem alterar o processo de integração.

ii. Base de dados integrada: esses produtos são fornecidos como a opção de compra do software de banco de dados e algumas das funcionalidades são incorporadas no banco de dados e não estão disponíveis separadamente na própria ferramenta ETL.

iii. Inteligência de negócios integrada: estes são os produtos do mesmo fornecedor que o software de BI. Em muitos casos, esses produtos são separados e o fornecedor afirmará que eles podem ser usados independentemente da ferramenta de BI.

iv. Produto de nicho: estes são os produtos que não se encaixam perfeitamente em nenhum dos grupos acima mencionados, mas ainda possuem funcionalidade ETL considerável neles.

As ferramentas ETL selecionadas para a comparação são líderes de mercado, embora existam tantas ferramentas no mercado (BEYER, et al., 2017). Para comparar essas ferramentas, os critérios ou bases em que estes são comparados devem ser universais (MADSEN, 2004).

Table7. Vários critérios para a revisão comparativa

Seq. Nº	Crítérios
1	Vendido em
2	Sozinho ou Integrado
3	Plataformas
4	Versão Analisada
5	Enginebased ou codegenerated
6	SaaS
7	Facilidade de uso
8	Reusabilidade
9	Depuração
10	Correção de erros de sintaxe
11	Compilador/Validador
12	Módulos Separados
13	Mecanismos de Dados
14	Junta tabelas como fonte

15	Suporte a informação de endereços
16	Conexões nativas
17	Conexões em tempo real
18	Agendamento
19	Pivoting/de-pivoting
20	SMP
21	MPP
22	Grid
23	Particionamento
24	Suporte CWM
25	Integração de Batch em tempo real
26	Pacotes / Aplicações empresariais

Fonte: o autor. Baseado (BEYER, et al., 2017; PassionnedGroup, 2017)

Tabela 8. Os produtos \ ferramentas analisados

Seq. Nº	Produto	Nome do Produto	Versão Analisada
1	IBM	Information Server	11.5
2	Informática	PowerCenter	10
3	Talend	Talend Open Studio for Data Integração	6.5
4	Oracle	Data Integrator (ODI)	12c
5	Microsoft	SQL Server Integrated Services	2017
6	SAS	Data Management	9.4
7	Kettle	Pentaho Data Integration	7.1
8	CloverETL	CloverETL	4.7

Fonte: o autor. Baseado (BEYER, et al., 2017; PassionnedGroup, 2017)

Análise comparativa das ferramentas de ETL conforme tabela 9.

Tabela 9. Análise comparativa das ferramentas de ETL analisadas

Crítérios	IBM Information Server	Informática PowerCenter	Talend Open Studio	Oracle Data Integrator	Microsoft SQL Server Integration Services	SAS Data Integration Studio	Kettle Pentaho	Clover ETL
Vendido em	1996	1996	2007	1999	1997	1996	2006	2005
Sozinho ou Integrado	Sozinho	Sozinho	Sozinho	Sozinho	Sozinho	Sozinho	Sozinho	Sozinho
Plataformas	6	5	7	6	1	8	4	7
Versão Analisada	8.1	9.5	5.2	11.1.1.5	10	v4.21	3.2	2.9.2
Engine based ou code generated	Ambos	engine based	code generated	code generated	Ambos	code generated	code generated	engine based
SaaS	Sim	Sim	Não	Sim	-	Não	Não sozinho	Não
Fácil de usar	Ordens lógicas	Sim	Sim	Amigável ao Usuário	Altamente	Altamente	Não	Não
Reusabilidade	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Debugging	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não
Correção de erros de sintaxe	Sim	Metade	Sim	Sim	-	Sim	Não	Sim
Compilador/Validador	Sim	Metade	Sim	Metade	Sim	Sim	Sim	Sim
Módulos Separados	Não	Sim	Sim	Não	-	Sim	Não	Sim
Mecanismos de Dados	Logging + triggers	Logging	Fila de mensagens + triggers	Fila de mensagens + Logging + triggers	Fila de mensagens + Logging + triggers	Fila de mensagens	Não	Fila de mensagens + triggers
Junta tabelas como fonte	Sim	Não	Sim	Sim	Não	Sim	Não	Não
Suporte a informação de endereços	Todos	Todos	Todos	Todos	-	-	Terceiros	Todos
Conexões nativas	41	50	35	22	4	18	20	7
Conexões em tempo real	2	6	3	3	2	3	3	3
Agendamento	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Pivoting/de-pivoting	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
SMP	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
MPP	Não	Não	Sim	Não	Não	Sim	Sim	Sim
Grid	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim
Particionamento	Sim	Sim	Não	Não	Sim	Sim	Sim	Sim
CWM suporte	Metade	Sim	Sim	Sim	Não	Sim	Não	Sim
Integração de Batch em tempo real	Sim	Sim	Sim	Sim	Não	Sim	Metade	Sim
Pacotes / Aplicações empresariais	8	7	9	8	1	5	2	0

Fonte: o autor. Baseado (BEYER, et al.,2017; PassionnedGroup, 2017)

4.1. Descrição e comparação dos critérios de análise

A comparação dos critérios mencionados na Tabela 9, e uma breve descrição geral das ferramentas, trazem elementos sobre as ferramentas e seus pontos fortes. Esta comparação apóia no sentido de identificar questões particulares inerentes a cada ferramenta.

Como pode ser visto na Tabela 7, as empresas cujas vendas começaram em 1996 como o Centro de Informação pela IBM, para algumas das empresas mais recentes, como a Talend Open Studio foram incluídos na análise comparativa (BEYER, et al.,2017). Apesar de todas as ferramentas de ETL acima serem ferramentas independentes no entanto elas têm suas próprias questões de pontuação. A análise comparativa para cada critério é dada a seguir (MADSEN, 2004):

Plataformas: Este critério significa quantas plataformas são suportadas pelo produto ETL, por exemplo, Windows (todas as versões foram contadas como um), Linux, Solaris, etc. Como pode ser visto *Microsoft SQL Server* tem o menor suporte à plataforma quanto o Windows enquanto o *SAS Data Management* fornece suporte para 8 tipos diferentes de plataformas que é um ponto positivo.

Enginebased ou codegenerated: Enquanto IBM Information Center e Microsoft SQL Server são ambos engine base ou codegenerated, todos os outros são baseados em um ou outro.

SaaS: Este critério identifica se o produto está disponível no modelo de software como um serviço e foi encontrado em: IBM Information Center, Informática PowerCentre e Oracle Data Integrator (ODI) que oferecem esta facilidade e os outros não. Isso significa que estes produtos podem ser oferecidos com serviços de computação, uma das mais recentes facilidades oferecidas pelas organizações. Este é um importante ponto positivo para esses produtos e talvez seja um dos motivos pelos quais eles são tão amplamente usados hoje.

Facilidade de uso: a facilidade de uso inclui o quão fácil é usar o produto, a rapidez com que pode ser aprendido, o número de dias de treinamento necessários para o desenvolvedor e o usuário para aprenderem o produto, design de elementos de tela, interface GUI e o mais importante é que funciona da maneira como a ETL deve funcionar. O Oracle Data Integrator (ODI) foi considerado o mais amigável, seguido do SQL Server e SAS Data Management. No entanto, isso não significa que os outros não sejam fáceis de usar é que essas três ferramentas e o Oracle Data Integrator (ODI), em particular, estão em conformidade com o critério acima mencionado mais do que os outros.

Reusabilidade: como os componentes são reutilizados se eles são orientados por parâmetros, ele suporta funções definidas pelo usuário para estar disponível para outros programas. Todas as ferramentas estão de acordo com este critério.

Depuração: excluindo o Pentaho e o CloverETL, todos os outros fornecem uma boa facilidade de depuração passo a passo ou linha a linha.

Correção de erros de sintaxe e nomes de campos: Pentaho, SQL Server e Informática não fornecem sugestão automática se houver um erro na sintaxe ou nomes de campo, enquanto isso está disponível em todas as outras ferramentas.

Compilador / Validador: é fácil localizar erros e se algum deles estiverem destacados no código com um clique é localizável e ajustável. Esta facilidade está disponível com todas as ferramentas.

Módulos separados: geralmente a ferramenta é composta por pelo menos dois módulos o módulo em tempo real e o módulo de lote. Eles podem ser comprados separadamente nas ferramentas Informática, Talend, SAS e CloverETL obtiveram esta provisão, enquanto este não é o caso da Oracle Data Integrator (ODI), IBM Information Center, SQL Server e Pentaho.

Mecanismo de dados: os dados mudam quando extraídos e transformados. Assim, a questão é como é reconhecido, isto é, como os dados alterados são reconhecidos. O IBM Information Center usa triggers e logs e entradas de diário para reconhecer os dados alterados, enquanto a Informática o faz com apenas os logs e entradas. Talend e CloverETL fazem isso com a fila de mensagens e banco de dados, o Oracle Data Integrator (ODI) e o SQL Server não deixam nenhuma opção para negligenciar tais mudanças, pois incorpora todas as três técnicas. O Pentaho destaca-se neste porque não fornece esta facilidade.

Junta tabelas como fonte: é possível juntar duas tabelas de maneira gráfica, deixando o banco de dados executar a união para não deixar a ferramenta ETL juntar as tabelas. InformaticaPowerCenter, SQL Server Integration, Pentaho e CloverETL não fornecem isso que é uma grande desvantagem.

Suporte a informação de endereços: todos os tipos de informações de endereço são suportados por todas as ferramentas.

Conexões nativas: quantas e quais conexões nativas a ferramenta ETL é compatível? (ODBC, OLE DB, JDBC e arquivos). InformaticaPowerCenter fornece o maior número de conexões nativas para as várias fontes de banco de dados, assim, a extração dessas fontes torna-se muito mais eficiente. O IBM Information Center e o Talend Open Studio também possuem muitas conexões nativas e seguem a Informática muito de perto. O SQL Server fornece apenas quatro tipos de conexões nativas.

Conexões em tempo real: quantos e quais tipos de produtos de filas de mensagens podem se conectar à ferramenta? Aqui também o Informatica Power Center possui o maior número de conexões, fornecendo as conexões máximas.

Agendamento: habilidade na ferramenta para agendar empregos com base em interdependências, em outras palavras, é o programador capaz de lidar com dependências. Todas as ferramentas hoje em dia suportam a funcionalidade de programação porque é considerada como uma necessidade básica para o ETL. Todas as ferramentas suportam o Agendamento.

Pivoting / de-pivoting: é possível transformar dados desnormalizados, colocar dados nos nomes das colunas, em linhas e ao contrário, transformar (altamente) dados normalizados para dados desnormalizados, colocando dados nas colunas. Mais uma vez esta facilidade está disponível em todas as ferramentas.

SMP: o Symmetric Multiprocessing é suportado? Padrão no Windows NT e no UNIX. Os processadores em sistemas SMP compartilham sua memória interna e externa. O SMP está disponível em todas as ferramentas mencionadas.

MPP: Todo processador em um sistema MPP possui sua própria memória interna e externa e banco de dados, permitindo que o alto desempenho seja alcançado. Esses bancos de dados devem ser sincronizados. A partir das ferramentas ETL retomadas para a revisão, apenas Talend, SAS, Pentaho e CloverETL possuem essa funcionalidade.

Grade: Um processo ETL pode ser executado em uma "grade" de computadores ou servidores? Somente os Serviços Integrados do SQL Server não fornecem a facilidade da grade enquanto todos os outros o fazem.

Particionamento: É possível particionar com base, por exemplo, em códigos de produtos, para determinar em qual máquina ou processador os dados devem ser processados. Talend e Oracle não permitem que a partição ocorra, todos os outros permitem.

Suporte CWM: a ferramenta ETL compatível com CWM, em outras palavras, é compatível com o Meta Model Warehouse Comum, ou um meta-modelo de armazenamento comum, o IBM Information Server, os Serviços Integrados do SQL Server e o Pentaho não fornecem essa facilidade.

Integração Batch (em lote) em tempo real: é possível definir dentro dos fluxos de processo da ferramenta ETL movendo e transformando dados em tempo real e em lote. Enquanto os Serviços Integrados do SQL Server não fornecem isso e o Pentaho fornece isso até certo ponto, todos os outros fornecem suporte total em integração em lote em tempo real.

Pacote / Aplicações empresariais: quantos pacotes / aplicativos corporativos a ferramenta pode ler meta-dados com um clique do mouse (por exemplo, SAP, Siebel, Peoplesoft, JD Edwards, Baan). Talend já incorpora 9 aplicações, seguido pelo IBM Information Server e Oracle Data Integrator (ODI) que podem ler 8, a Informatica PowerCenter pode ler 7 e o CloverETL que não tem incorporada esta funcionalidade.

4.2. Descrição das ferramentas avaliadas

Todas as ferramentas possuem a disponibilidade e flexibilidade da implantação utilizando ETL. O poder da ferramenta, inicialmente imputado para a geração de dados para um data warehouse, traz habilidades que auxiliam, agilizam e tornam seguras as conversões de dados tão necessárias na integração e troca de softwares e aplicativos.

A seguir a explanação sobre algumas das aplicações analisadas:

IBM Information Server

Fornecer uma grande flexibilidade e é direcionado ao mercado com uma visão em mente de plataforma comum de metadados. O Information Server fornece alto nível de satisfação dos clientes e uma variedade de iniciativas. Embora seja fácil de usar, torna-se muito pesado porque os dados envolvidos estão em GBs e a versão 11.x requer muito poder de processamento (LARSON, 2016).

InformaticaPowerCenter

Oferece uma tecnologia muito sólida, uma curva de aprendizado direta, a capacidade de abordar esquemas de integração de dados em tempo real e é altamente especializada em ETL e Integração de Dados. Tem um histórico consistente com tamanho e recursos substanciais no mercado de fornecedores de ferramentas de integração de dados.

Talend

É uma ferramenta de integração de dados de código aberto, mas não um BI completo. Ele usa uma abordagem de geração de código. Usa uma GUI. Possui recursos de qualidade de dados: a partir de sua própria GUI, escrevendo mais consultas SQL personalizadas e Java (TALEND, 2017).

Microsoft SQL Server Integration Services

O SQL Server Integration Services (SSIS) fornece facilidade e velocidade de implementação com recursos padronizados de integração de dados, em tempo real, baseados em mensagens, que são de custo relativamente baixo e oferecem um excelente suporte e modelo de distribuição. No entanto, ele não suporta ambientes não-Windows (LARSON, 2016).

Oracle Data Integrator (ODI)

Não há dúvida de por que está sendo considerado como um dos líderes nos mercados ETL, isto porque possui uma conexão ajustada com todos os aplicativos de armazenamento de dados Oracle e a tendência de integrar todas as ferramentas em uma única aplicação e em um ambiente (ORACLE, 2017).

SAS Data Management

Fornece um excelente suporte e, acima de tudo, a ferramenta de integração de dados é muito poderosa e com muitos recursos de gerenciamento múltiplo. É um grande apoio para empresas de classe empresarial também para aqueles que são médios e menores. Pode funcionar em muitos sistemas operacionais e reunir dados através de várias fontes, muito flexível (SAS, 2017).

Pentaho (Kettle)

Pentaho é uma suíte de BI de código aberto comercial que possui um produto chamado Kettle para integração de dados. Ele usa uma abordagem inovadora meta-driven e possui uma GUI forte e muito fácil de usar. Possui um mecanismo java autônomo que processa os trabalhos e tarefas para mover dados entre vários bancos de dados e arquivos diferentes (PENTHO, 2017).

CloverETL

Fornece integração de dados, automação de fluxo de trabalho através de fluxos de trabalho. Pode transformar os dados à vontade. É uma ferramenta visual que substitui scripts todos os dias e fornece controle total dos fluxos e processos de dados (CloverETL, 2017).

Conclusão sobre as ferramentas analisadas

Nesta comparação de ferramentas de software ETL, é possível observar que o *Oracle Data Integrator (ODI)* e o *IBM Information Servers* são aqueles que satisfazem as necessidades das grandes empresas e as outras ferramentas mencionadas, têm seus próprios aspectos a serem implementados.

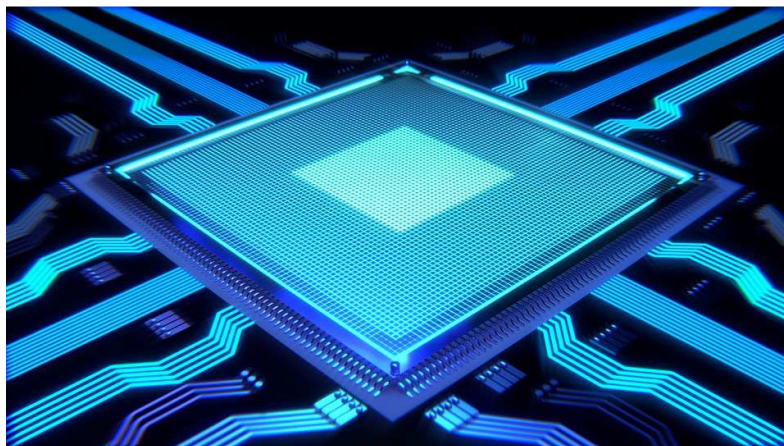
A maioria das ferramentas mais novas podem ser mais lentas e com menos funções para satisfazer as necessidades das organizações maiores, mas podem ser aproveitadas para empresas medianas ou pequenas, onde a performance atenderia perfeitamente.

Alguns dos aspectos deixados de fora e não incluídos nos critérios, assim como o preço, pois cada organização só pode decidir comprar o produto, avaliando as funcionalidades e os benefícios que se espera que sejam colhidas. Além disso, nem todos os produtos garantem fornecer todas as funcionalidades.

Há também a questão do compromisso quanto ao que deve funcionar para uma empresa, e que pode não funcionar para outra empresa. Por isso, embora apresenta a revisão e avaliação sobre essas ferramentas, as empresas podem decidir ainda, atribuindo um peso aos critérios e às variáveis externas que as empresas sentem ser importantes e, em seguida, fazer a decisão pela ferramenta ideal.

A ETL é apenas uma parte da suíte de BI oferecida pelos fornecedores. O preço do produto não foi incluído porque a maioria das organizações não revela seus preços para se manterem competitivas e muitas soluções são adaptadas às necessidades específicas de uma organização por meio de cotações.

5. Query cache



Disponível em: <https://pixabay.com/photo-2217771/>

Para o acesso rápido ao banco de dados, é recomendável utilizar o cache da consulta. O cache da consulta armazenará toda a gravação da consulta executada. O cache da consulta manterá registro de consultas recém-executadas. O objetivo principal do cache de consulta é reduzir o tempo de resposta da consulta. Aumentará a capacidade intelectual do data warehouse para que o sistema memorize o último trabalho que realizou.

Esta memória é utilizada posteriormente para responder ao resultado das consultas que foram executadas anteriormente pelos usuários. O cache manterá dois estados válidos e estado inválido.

Quando qualquer consulta é enviada pelo usuário, primeiramente é examinada a memória cache, para verificar se a consulta solicitada já está armazenada no cache, se a consulta estiver armazenada, verifique se o estado é válido ou inválido. Se o estado for válido, os dados podem ser acessados e se o estado for inválido, os dados não podem ser acessados.

Se o usuário enviar uma consulta de inserir, atualizar, excluir ou de drop, os dados serão alterados no banco de dados e o estado da consulta relacionada será inválido. Agora Dados e consulta de estado inválidos não podem ser acessados pelo usuário. Isso pode economizar tempo importante e melhorar o

desempenho do armazenamento de dados ao não reavaliar as consultas que já estão armazenadas no cache.

Um analista coloca uma consulta para apresentar o emprego de uma empresa, que trabalha sob o gerente_id é 100,101,201. A consulta será semelhante à seguinte:

```
SELECTEmp_id, name, salary, manager_id  
FROMemployees  
WHEREmanager_id IN (100, 101, 201);
```

Quando a consulta é enviada, o cache da consulta será examinado para verificar se esta consulta está disponível ou não e se o estado é válido ou inválido. Se não estiver disponível, a consulta será avaliada e o resultado será armazenado no cache da consulta. Os resultados da consulta são mostrados na tabela 7.

Tabela 10 - Saída da Query

Emp_id	Name	Salary	Manager_id
202	Mukesh	6000	201
200	Mohan	4400	101
205	Sohan	12000	101
101	Rohit	17000	100
102	Sanjay	14000	100

Se algum outro usuário enviou a mesma consulta, o resultado será recuperado do cache de consulta porque essa consulta já está armazenada no cache.

Supondo que outro usuário queira o empregado de uma empresa, cujo salário maior seja igual a 10000 e o gerente seja manager_id igual a 100, 101. A consulta será semelhante à seguinte:

```
SELECTEmp_id, name, salary, manager_id  
FROMemployees  
ONDE manager_id IN (100, 101) and salary >= 10000;
```

Quando a consulta é enviada, a memória cache é examinada. A mesma consulta é armazenada na memória cache e o estado é válido, então podemos obter o resultado da segunda query como mostrado na tabela 8.

Tabela 11 - Saída da nova query

Emp_id	Name	Salary	Manager_id
205	Sohan	12000	101
101	Rohit	17000	100
102	Sanjay	14000	100

O resultado da segunda Query será gerado a partir do conjunto de resultados da primeira Query em vez de passar por todos os dados armazenados no data warehouse. Este processo economizará muito tempo e esforço necessário para passar por todos os registros.

A consulta SELECT, o cache primeiro particiona a consulta para acessar todas as tabelas particionadas. Em seguida, verifica se cada consulta particionada é previamente armazenada em cache. Caso contrário, então, verifica se qualquer consulta previamente armazenada em cache pode fornecer uma resposta parcial.

Se tal consulta existe e o estado é válido, ele envia uma consulta restante ao banco de dados. Quando o banco de dados retorna o resultado para a consulta restante, ele mescla o resultado com o resultado parcial obtido do cache. Finalmente, ele combina os resultados de todas as consultas particionadas e retorna o resultado esperado ao cliente.

Quando uma atualização ou consulta de exclusão é recebida, é encaminhada para o banco de dados. Quando a resposta do banco de dados é recebida, o cache invalida todas as entradas de cache

Dependendo das tabelas afetadas ou das colunas afetadas. Essas referências são então excluídas da estrutura de dados do esquema do banco de dados.

A situação do cache da consulta é observada classificando em:

- Inválida - se a consulta não for armazenada no cache, o estado será inválido. Se os dados forem atualizados pelo usuário por qualquer inserção de consulta, a exclusão de atualização do estado será inválida.
- Válida - se a consulta for armazenada no cache e não atualizada no banco de dados de qualquer inserção de consulta, a atualização será excluída. O estado será válido.

Nosso problema é que temos uma consulta e um resultado de consulta armazenados no cache. Mas se o banco de dados for atualizado com os novos dados, o resultado da consulta do cache refletirá em dados antigos. Criaremos um mecanismo de estado;

A primeira query é enviada pelo usuário e seu resultado é armazenado no cache da consulta. Quando o próximo usuário envia a mesma query ao banco de dados atualizado, o cache da consulta verificará o estado, se o estado for inválido, significa que o banco de dados foi atualizado com novos dados. Agora, a consulta não precisa passar por todos os registros. Ele obterá o último índice do resultado da consulta armazenado no cache da consulta. Em seguida, começará a pesquisar os registros que atendem aos critérios de consulta a partir desse índice. Isso pode economizar muito tempo e esforço necessários para pesquisar a grande quantidade de dados.

6. Considerações finais

Os negócios de hoje obrigam as empresas a operar de forma diferenciada, e coexistem através de sistemas de informação que necessitam de integração.

As empresas de armazenamento de dados têm um dilema de escolher o processo de Extração, Transformação e Leitura e tem opções de ferramentas ETL que podem levar a uma série de perdas tanto monetariamente pelo tempo, quanto pela quantidade de trabalho laborioso que os trabalhadores dedicaram.

A equipe da organização envolvida na ETL pode escolher entre uma variedade de ferramentas, mas necessitam explorar ou ter o conhecimento de suas características para resultar em um bom processo de tomada de decisão.

É preciso realizar uma análise comparativa das principais ferramentas de ETL só para familiarizar os usuários com as suas características e também suas desvantagens.

Outro ponto importante é que as funções ETL precisam ser realizadas por uma equipe ETL competente e treinada. A equipe ETL é liderada por um especialista ETL. É responsabilidade do arquiteto ETL elaborar um processo ETL abrangente e eficaz para efetuar o carregamento de um data warehouse ou uma exportação que se fizer necessária.

O arquiteto / especialista ETL garante que os processos ETL tenham força e consistência. O arquiteto ETL trabalha em estreita coordenação com os usuários empresariais e identifica quais dados e em que níveis de detalhes são necessários.

A materialização da visão é uma estratégia usada para fornecer respostas rápidas às consultas dos usuários.

Sem as ferramentas ETL, os relatórios analíticos não são possíveis e o ETL em si é a base necessária de qualquer ferramenta de Business Intelligence (BI) usada por uma organização. Uma vez que, a ETL envolve o trabalho em três partes de extrair dados das fontes, transformando-os em um formato unificado e, finalmente, carregando-os unificados em um data warehouse consistente.

A utilização do método de memória cache para melhorar o desempenho e a velocidade nos processos de ETL, minimizam significativamente o tempo de resposta. O principal objetivo desta técnica é armazenar consultas e seus resultados correspondentes. Se uma consulta semelhante for enviada por qualquer outro usuário, o resultado será obtido usando a memória cache.

A técnica Query Cache é armazenar consultas e seus resultados correspondentes. Se uma consulta semelhante for enviada por qualquer outro usuário, o resultado será obtido usando a memória cache.

A informação de consultas anteriores é usada para gerar resultados. Finalmente, os resultados são eficientemente fundidos para alto desempenho. Este método ativo melhora bastante o desempenho e velocidade do ETL.

Elementos Complementares

Título: Pentaho KettleSolutions: Building Open Source ETL Solutionswith Pentaho Data Integration



Autor: Matt Casters, Roland Bouman, Jos van Dongen

Editora: JOHN WILEY CONSUMER

Sinopse: Um guia completo para o Pentaho Kettle, o conjunto de ferramentas Pentaho Data Integration para ETL. Este livro prático é um guia completo para instalar, configurar e gerenciar Pentaho Kettle. Se você for um administrador de banco de dados ou

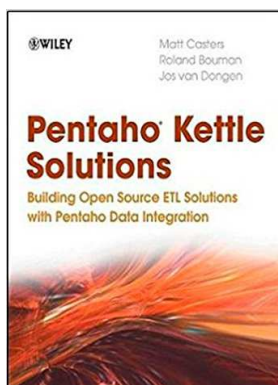
um desenvolvedor, você primeiro se informará sobre os conceitos básicos da ferramenta Kettle e sobre como aplicá-la para criar soluções ETL antes de avançar para conceitos especializados, como os modelos de agrupamento, extensibilidade e segurança de dados. Saiba como desenhar e construir cada fase de uma solução ETL.

Título: Kettle - Pentaho Data Integration

Autores: Adrián Sergio Pulvirenti e María Carina Roldán

Editora: PacktPublishing

Sinopse: Este livro tem o passo-a-passo para resolver os problemas de manipulação de dados (ETLs) usando PDI na forma de receitas. Tem muita bem organizada dicas, screenshots, tabelas e exemplos para auxiliar a compreensão rápida e fácil.



Referências

BEYER, Mark A.; THOO, Eric; SELVAGE, Mei Yang; ZAIDI, Ehtisham. **Magic Quadrant for Data Integration Tools**. Stamford, USA: Gartner, 2017.

Disponível em: <https://www.gartner.com/doc/reprints?id=1-48R9Q5U&ct=170803&st=sg>. Acessado em: Agosto/2017.

CloverETL. **CloverETLProducts**. Acessado em Junho/2017. Disponível em: <http://www.cloveretl.com/products>.

ELMASRI, Ramez, NAVATHE, Shamkant B. **Fundamentals of database systems**. Boston: Pearson, 6a. Edição, 2011.

GONÇALVES, Rodrigo Ribeiro. **Integração de Dados na Prática**. São Paulo: Érica, 2012.

GOUR, Vishal. **Strategy to make superior Data warehouse**. International Conference on advance computing and creating entrepreneurs. 2010, 139-143.

LARSON, Brian. **Delivering Business Intelligence with Microsoft SQL Server**. 4a. Edição. New York: McGraw-Hill, 2016.

MADSEN, Mark. **Criteria for ETL Product Selection**. Information Management. Disponível em: <http://www.information-management.com/infodirect/20041001/1011217-1.html?pg=1>, 2004. Acessado em: Agosto/2017.

NUTT, W.; SAGIV, Y.; SHURIN, S. **Deciding Equivalence among Aggregate Queries**, 17th Symposium on Principles of Database Systems (PODS'98, Seattle, Washington, USA, June 1-3), 1998.

Oracle. **Oracle Data Integrator (ODI)**. Acessado em: Julho/2017. Disponível em: <http://www.oracle.com/technetwork/middleware/data-integrator/overview/index.html>

PassionnedGroup. **ETL Tools & Data Integration Survey**. PassionnedGroup, 2017. Disponível em: <https://www.etltool.com/kb/etl-tools-data-integration-survey/>, acessado em: Agosto/2017.

PassionnedGroup. **List of ETL tools**. PassionnedGroup, 2017. Disponível em: <https://www.etltool.com/list-of-etl-tools/>, acessado em: Agosto/2017.

Pentaho. **Pentaho Data Integration (Kettle)**. Acessado em: Julho/2017. Disponível em: <http://kettle.pentaho.com/>

SAS Data Integration. **Studio SAS Products & Solutions**. Acessado em: Julho/2017. Disponível em: <http://support.sas.com/software/products/etls/>

SATAPATHY, Ahimanikya. **Building an ETL Tool**, SOA/Business Integration, Sun Microsystems, 2006.

SHIM J.; SCHEUERMANN, P.; VINGRALEK, R. **DynamicCachingof Query Results for DecisionSupport Systems**, Proceedings of the 11th International Conference on Scientific and Statistical Database Management (SSDBM'99, Cleveland, Ohio, USA, July 28-30), 1999.

SULTAN, Fahad; AZIZ, Abdul. **Ideal Strategy to Improve Datawarehouse Performance**. (IJCSE) International Journal on Computer Science and Engineering. Vol.02, No.02, 2010, 409-415.

SWEIGER, Mark; MADSEN, Mark R.; LANGSTON, Jimmy; LOMBARD, Howard; **Clickstream Data Warehousing**. Universidade da Califórnia: John Wiley & Sons, 2002.

Talend. **Talend Open Studio**. Acessado em: Julho/2017. Disponível em: <https://www.talend.com/products/data-integration/>

ZODE, Madhu. **The Evolution of ETL**. Acessado em: Julho/2017. Disponível em: <http://hosteddocs.ittoolbox.com/mz071807b.pdf>