



# DATA ANALYTICS (DW/DM/BI) E BIG DATA

Prof. Msc. Marcos Alexandruk

[alexandruk@uni9.pro.br](mailto:alexandruk@uni9.pro.br)

<https://github.com/alexandruk/dw>



## Aula 01

### Apresentação da disciplina

## Introdução

Data Warehouse se tornou uma realidade em muitas empresas e a cada dia mais empresas adotam sistemas de **Inteligências de Negócios** para tomada de decisões, visando aumentar o lucro.

Com o crescimento da demanda por este tipo de tecnologia, cresce também a necessidade da formação de **novos profissionais** capacitados a suportar e desenvolver este segmento, então vamos estudar.

## Introdução

Estamos vivenciando a era do conteúdo, em que aumentar o capital intelectual de uma empresa é uma necessidade competitiva. As organizações que usam, de forma eficiente, a tecnologia da informação adquirem conhecimento e velocidade para alcançar um diferencial nos mercados em que atuam e se tornam capazes de alavancar seus negócios.

## Introdução

O ambiente de dados para suportar os processos gerenciais e de tomada de decisão é diferente do ambiente convencional de processamento de transações.

Data Warehouse tem o objetivo de integrar as informações dos principais processos empresariais e possui a capacidade de consolidar tais dados, adquiridos de diferentes acervos, para fins de exploração e análise, ampliando o conteúdo informacional a fim de atender às expectativas e necessidades de nível gerencial e estratégico na empresa.

## Introdução

As empresas dependem de sua capacidade de analisar, planejar e reagir, de forma rápida e imediata, às mudanças nas condições de seus negócios. Para que isso aconteça, é necessário que a organização disponha de informações em quantidade e qualidade.

## Introdução

Diariamente, dados sobre os mais variados aspectos dos negócios da empresa são gerados e armazenados, e passam a fazer parte dos recursos de informação dessa empresa.

Entretanto, tais informações encontram-se, em geral, espalhadas em diversos sistemas e exigem um esforço considerável de integração para que possam dar suporte efetivo à tomada de decisão de gerentes e executivos.

Por isso, estas decisões normalmente são tomadas com base na experiência dos administradores, quando poderiam também ser baseadas em fatos históricos que foram armazenados pelos diversos sistemas de informação utilizados pelas organizações.

## Introdução

Quando desenvolvemos um sistema transacional, seu propósito principal é suportar as transações de negócio da empresa, exemplificando: sistema de vendas, compras, controle de estoque, etc. Estes tipos de sistemas não são projetados para gerar e armazenar as informações estratégicas, por esse motivo, um novo conjunto de conceitos e ferramentas vem ganhando enorme destaque nos últimos anos: a tecnologia de Data Warehouse , que oferece às organizações uma maneira flexível e eficiente de obter as informações necessárias aos seus processos decisórios.



## Data Warehouse

As empresas realizam muitas perguntas que não conseguimos responder com sistemas transacionais, como:

- Como reduzir os custos sem impactar na qualidade do produto?
- Quem são os melhores clientes para vender seus produtos?
- Como aumentar a participação de seus produtos no mercado?
- Quais são os segmentos de mercado mais significativos para seu negócio?
- Como maximizar as vendas por cliente?
- Quais promoções realizar?
- Para quais produtos devo realizar promoções?
- Em que momento realizar as promoções?
- Qual o estoque mínimo necessário por produto em determinado período?
- Que produtos têm maior índice de retorno?
- Quais clientes possuem uma conta corrente, mas não uma conta poupança?
- Quais clientes possuem mais de um tipo de conta?
- Quem são os meus clientes mais rentáveis entre os que compram mais?
- Como classificar os clientes por faixas usando métricas analíticas e estatísticas avançadas?

Conhecer a resposta destas e outras perguntas é **vantagem competitiva** que toda empresa deseja, mas normalmente elas desconhecem o potencial dos dados de seu sistema de informação.

## Data Warehouse

O Data Warehouse nasceu a partir do reconhecimento da importância do valor da informação nas organizações. Ele é um ambiente expansível e planejado para a análise de dados imutáveis.

Tais dados são lógica e fisicamente transformados, provenientes de múltiplas fontes, atualizados e mantidos por um longo período de tempo, expressos em termos do negócio e resumidos para uma análise eficiente.

Assim, um Data Warehouse é uma plataforma com dados integrados e sua qualidade melhorada para apoiar as tomadas de Decisões pelos executivos das empresas.

## Data Warehouse

Em termos tecnológicos, um Data Warehouse é uma combinação de várias tecnologias, tendo como primeiro objetivo a integração efetiva de bases de dados operacionais em um ambiente que habilita o uso estratégico dos dados.

Estas tecnologias incluem sistemas gerenciadores de banco de dados relacional e multidimensional, modelagem de metadados e repositórios, interfaces gráficas para usuário, etc.

## Data Warehouse

### Algumas definições para Data Warehouse

- É um banco de dados orientado por assunto, integrado, não volátil e histórico, criado para suportar o processo de tomada de decisão, enquanto Data Mart é uma coleção de assuntos, organizados para dar suporte à tomada de decisão e baseados nas necessidades de um determinado departamento.
- É um conjunto de conceitos, métodos e recursos tecnológicos que habilitam a obtenção e distribuição de informações geradas a partir de dados operacionais, históricos e externos, visando proporcionar subsídios para tomadas de decisões gerenciais e estratégicas.
- É uma cópia dos dados da transação; especificamente estruturado para consulta e análise.
- É o processo de integração dos dados corporativos de uma empresa em um único repositório de dados a partir do qual os usuários finais podem facilmente executar consultas, gerar relatórios e fazer análises
- É o resultado de um processo de armazenagem de dados específicos e integrados, de fontes heterogêneas, para a realização de consultas e análises dimensionais.

## Data Warehouse

O Data Warehouse surgiu na década de 80 com destaque aos trabalhos de **Bill Inmon**, que tinha o objetivo de resolver a geração de informações empresariais, processo esse que é falho nos bancos de dados transacionais.

Mais conhecido como o "Pai do Data Warehouse", Bill Inmon tornou-se o autor mais conhecido em todo o mundo na área de Data Warehouse/Business Intelligence (mais de 50 livros e 650 artigos). Em 2007, Bill foi nomeado pela Computerworld como uma das "Dez pessoas que importavam de TI nos últimos 40 anos".

## Referências

CARVALHO, Luis A. V. **Datamining: A mineração de dados no marketing, medicina, economia, engenharia e administração.** 1. ed. Rio de Janeiro: Ciência Moderna, 2005.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Data Mining: um guia prático.** Rio de Janeiro: Campus, 2005.

MACHADO, Felipe Nery Rodrigues. **Tecnologia e projeto de data warehouse.** 2. ed. São Paulo: Érica, 2006.

SILBERSCHATZ, A.; KORTH, H.; SUBARSHAN, S. **Sistema de Banco de Dados.** 5. ed. Rio de Janeiro: Campus, 2006.

TAUB, Benjamin et al; **Oracle 8i data warehouse.** Rio de Janeiro: Campus, 2001.



## Aula 02

### Introdução e o Ambiente de um Data Warehouse

## Objetivo

Entender a evolução do desenvolvimento de um ambiente de banco de dados adequado à análise de negócios e ao apoio à tomada de decisões gerenciais e estratégicas.



## Introdução

Data Warehouse é o resultado de um processo de armazenagem de dados específicos e integrados, de fontes heterogêneas para a realização de consultas e análises dimensionais.

Data Warehouse representa um importante diferencial competitivo e não é um produto e sim um ambiente.

## Introdução

Uma organização moderna precisa de Sistemas de Informações eficientes e fáceis de se utilizar, a fim de sobreviver e obter sucesso em um ambiente globalizado e altamente competitivo. Neste contexto, pode-se citar várias razões para construir estes sistemas, que são:

- As decisões precisam ser tomadas rapidamente e corretamente, usando todos os dados disponíveis.
- Os usuários de sistemas de informações são especialistas de domínio de negócio, não profissionais de computação.
- O volume de dados dobra a cada 18 meses, o que afeta o tempo de resposta e, incontestavelmente, a habilidade em compreender seu conteúdo.
- A competição aumenta dia após dia nas áreas de inteligência empresarial, bem como o valor agregado de informações.

## Introdução

Com a evolução dos sistemas transacionais, as empresas passam a investir nas necessidades de gerar informações executivas (business intelligence).

Existe uma grande dificuldade dos desenvolvedores de sistemas para entender a diferença entre informação e dado.

Esta afirmação é a síntese da grande diferença entre um projeto de banco de dados para sistemas transacionais (OLTP) e um projeto de banco de dados para Data Warehouse, com tecnologia OLAP.

**OLTP - Online Transaction Processing**

**OLAP - Online Analytical Processing**

## Introdução

Os **dados** não possuem significado relevante e não conduzem a nenhuma compreensão. Representam algo que não tem sentido a princípio. Portanto, não podem ser usados para embasar conclusões, muito menos respaldar decisões.

A **informação** é a ordenação e organização dos dados de forma a transmitir significado e compreensão dentro de um determinado contexto. Representam o conjunto ou consolidação dos dados de forma a fundamentar o conhecimento.



## Ambiente de um Data Warehouse



grande verde casa



A casa verde é grande.

## Ambiente de um Data Warehouse

Vamos imaginar um exemplo de diagnóstico médico, analisando um exame de rotina:

Mede-se a temperatura a pressão arterial. Obtêm-se **dados** para uma futura análise de um especialista: o médico.

Se projetarmos um sistema para armazenar e manipular esses dados de várias formas, ainda não trataremos da informação em modo sistêmico. O que interessa saber é se o paciente: vai ter uma parada cardíaca ou um enfarto, se sua saúde está boa, etc.

Estas respostas são **informações** e podem ser projetadas através de gráficos de tendência ou cálculos estatísticos simulando possibilidades, desenvolvendo aplicações de sistema.

Para suprir esta necessidade de guardar dados para uso futuro, tem-se o surgimento do conceito de Armazém de Dados (Data Warehouse).

## Ambiente de um Data Warehouse

Nos dias de hoje, não podemos imaginar um médico realizando um diagnóstico sem ao menos nos pedir um exame, caso isto ocorresse, não confiaríamos em sua opinião. Então porque devemos confiar em um empresário que toma decisões com pouca informação ou nenhuma? Não devemos pensar em fazer negócios e muito menos comprar ações ou investir nosso dinheiro em empresas em que a tomada de decisão é feita na "sorte".

Quando vamos a um médico, ficamos satisfeito quando o diagnóstico é realizado após exames detalhados de nossas condições físicas (exames de sangue, tomografias, etc.).

Similarmente, os executivos necessitam, para diagnosticar e administrar as tendências de negócio, de um ambiente que lhes permitam executar exames de seus dados com a mesma capacidade, profundidade, transparência e evolução.

## Ambiente de um Data Warehouse

Quando um médico, além de seus exames, analisa seu histórico pessoal e familiar, com isto, pode diagnosticar tendências a um quadro de diabetes, problemas cardíacos, infecções respiratórias, etc.

É necessário que o executivo visualize suas tendências de vendas, por exemplo. A sazonalidade e a regionalização de vendas de determinados produtos, etc. Uma análise de dados históricos pode nos apresentar indicadores de crescimento ou sinalizadores para os negócios.



## Ambiente de um Data Warehouse

**Algumas diferenças entre dados de Sistemas Transacionais e Data Warehouse.**

Dados Transacionais (OLTP)	Dados em um Data Warehouse (OLAP)
Baseados em aplicações.	Baseados em assuntos ou negócios.
Detalhados.	Resumidos ou refinados.
Exatos em relação do momento de acesso.	Representam valores de momentos já decorridos ou instantâneos.
Acessados uma unidade por vez.	Acessados um conjunto por vez.
Alta disponibilidade.	Disponibilidade atenuada.
Não contemplam a redundância.	A redundância não pode ser ignorada.
Estrutura fixa: conteúdos variáveis.	Estrutura flexível.
Pequena quantidade de dados usada em um processo.	Grande quantidade de dados usada em um processo.
Atendem às necessidades cotidianas.	Atendem às necessidades gerenciais.
Alta probabilidade de acesso.	Baixa, ou modesta probabilidade de acesso.

## Ambiente de um Data Warehouse

Razões tecnológicas para a existência de Data Warehouse:

**Primeira:** o Data Warehouse é projetado para resolver a incompatibilidade de sistemas de informações transacionais e operacionais. Estas duas classes de sistemas são projetadas para satisfazer exigências diferentes, mas frequentemente incompatíveis.

**Segunda:** a infraestrutura (IT) muda rapidamente, do mesmo modo que suas capacidades aumentam. Isto pode ser evidenciado através dos seguintes pontos:

- O preço dos computadores que operam em uma velocidade medida em MIPS (milhões de instruções por segundo) continua caindo, enquanto o poder dos microprocessadores dobra a cada 2 anos.
- O preço de armazenamento digital está diminuindo.
- A banda passante das redes está aumentando, enquanto o preço de banda, diminuindo.
- O local de trabalho é crescentemente heterogêneo em termos de hardware e software.
- Os sistemas legados precisam, e podem, ser integrados com novas aplicações.

## Referências

CARVALHO, Luis A. V. **Datamining: A mineração de dados no marketing, medicina, economia, engenharia e administração.** 1. ed. Rio de Janeiro: Ciência Moderna, 2005.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Data Mining: um guia prático.** Rio de Janeiro: Campus, 2005.

MACHADO, Felipe Nery Rodrigues. **Tecnologia e projeto de data warehouse.** 2. ed. São Paulo: Érica, 2006.

SILBERSCHATZ, A.; KORTH, H.; SUBARSHAN, S. **Sistema de Banco de Dados.** 5. ed. Rio de Janeiro: Campus, 2006.

TAUB, Benjamin et al; **Oracle 8i data warehouse.** Rio de Janeiro: Campus, 2001.



## Aula 03

### Ecossistema de informações e Sistemas Transacionais (OLTP)

## Objetivo

Estudar o ciclo de vida das informações, desde o seu início até o momento que se torne obsoleto e descartado. Entender a importância dos sistemas transacionais na evolução do ambiente de banco de dados, adequado à análise de negócios apoiando a tomada de decisões gerenciais e estratégicas.

## Sistemas Transacionais (OLTP)

No mundo corporativo, a informação é um dos ativos (patrimônio) mais importantes, residente na cabeça de seus empregados e muitas vezes supera o valor contábil da empresa.

## Sistemas Transacionais (OLTP)

Estamos vivendo na era de "sobrecarga de informações". As companhias possuem diversos sistemas, muitas vezes com tecnologias diferentes, utilizados para automatizar e administrar seus recursos.

Estes sistemas geram e retém uma enorme quantidade de dados estruturados de várias maneiras, além disso, as empresas também armazenam grandes volumes de dados não estruturados na forma de e-mail, documentos (doc, xls, pdf, etc.) e imagens.

As informações estruturadas e não estruturadas devem ser armazenadas e retidas de dentro destes sistemas.

## Sistemas Transacionais (OLTP)

O ciclo de vida da informação passa pelas fases de:

- criação ou captura;
- armazenamento;
- versionamento;
- indexação;
- gestão;
- limpeza;
- distribuição;
- publicação;
- pesquisa;
- arquivo (ou descarte).



## Sistemas Transacionais (OLTP)

Os sistemas estão provocando mudanças organizacionais e administrativas, trazendo grandes desafios para administração. Hoje, qualquer sistema que colocamos dentro da empresa deve se integrar aos já existentes e participar do sistema de informações que atende aos diversos níveis e funções organizacionais.

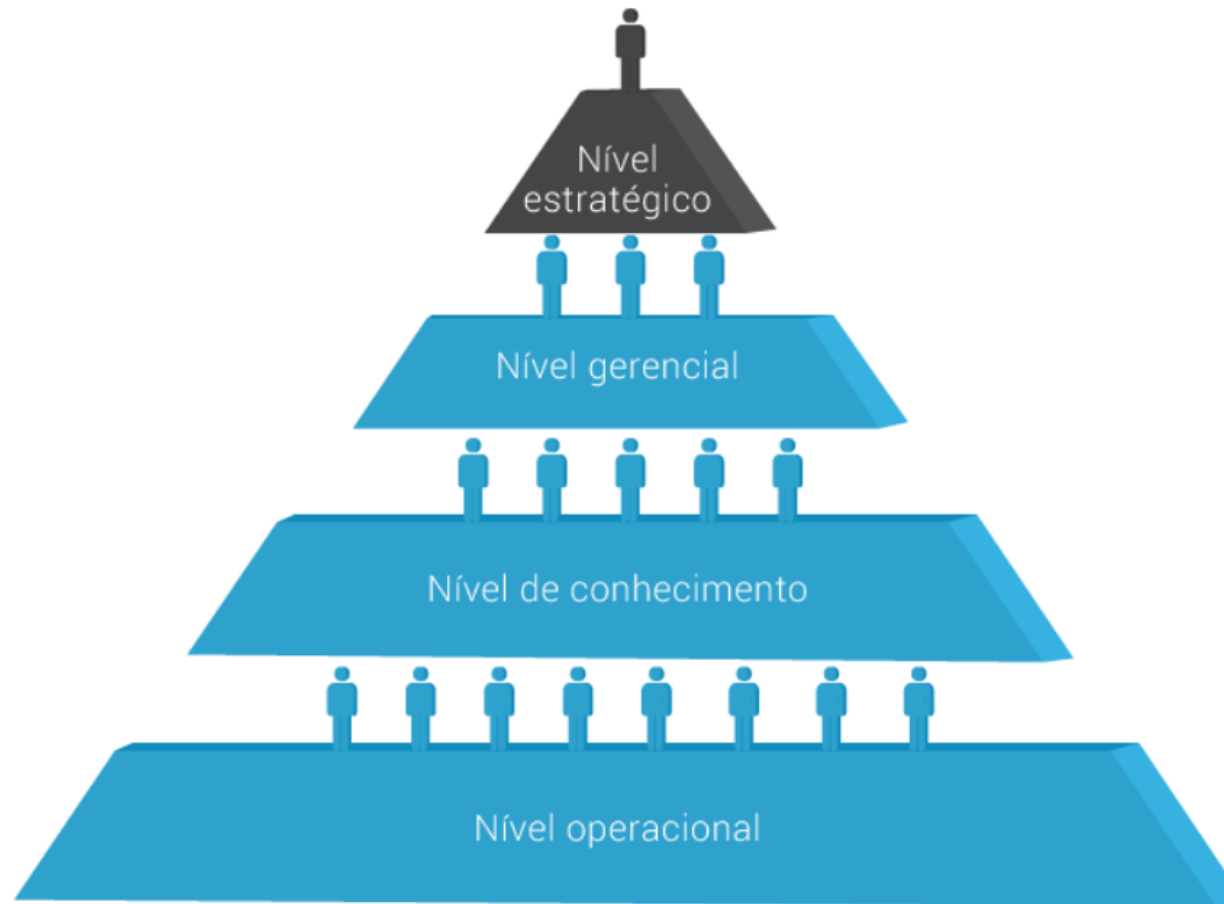
## Sistemas Transacionais (OLTP)

Devido à existência de diferentes interesses e especialidades em cada nível organizacional, são necessários diversos tipos informações, pois nenhum sistema individual pode atender a todas as necessidades de uma empresa. Vamos dividir os tipos de sistemas em quatro níveis organizacionais:

- **Sistemas do nível operacional:** suporte a gerentes operacionais em transações como vendas, contas a receber, depósitos, fluxo de matéria-prima, etc.
- **Sistemas do nível de conhecimento:** envolvem as estações de trabalho e automação de escritório a fim de controlar o fluxo de documentos.
- **Sistemas do nível gerencial:** atendem a atividades de monitoração, controle, tomadas de decisões e procedimentos administrativos dos gerentes médios.
- **Sistemas de nível estratégico:** que ajudam a alta gerência a enfrentar questões e tendências, tanto no ambiente externo como interno a empresa.

Existem sistemas de informações desenvolvidos com capacidade de atender mais de um nível organizacional, mas, na maioria dos casos, desenvolvemos sistemas de informações focados em um nível organizacional e o potencializamos com integrações com outros sistemas.

## Sistemas Transacionais (OLTP)



## Sistemas de Processamento de Transações (Operacional)

Devem ser integrados e atendem ao nível operacional, realizando transações rotineiras, como folha de pagamento, pedidos, vendas, controle de estoque, etc. Através deles, os gerentes monitoram operações internas e externas da empresa. São críticos, pois se deixarem de funcionar, podem causar prejuízos a outras empresas e a própria.

## Sistemas de Conhecimento

São de automação de escritório, atendem às necessidades do nível de conhecimento envolvendo duas classes de trabalhadores: a primeira com "formação universitária" (?!), como engenheiros, analistas financeiros, analistas administrativo etc., e a segunda com trabalhadores de dados, como secretárias, técnicos administrativos, contadores, arquivistas etc. As duas classes se diferenciam, pois trabalhadores de conhecimento criam informações e trabalhadores de dados manipulam, usam informações prontas, a produtividade dos últimos é aumentada com o uso destes sistemas, que coordenam e comunicam diversas unidades.

## Sistemas de Informação Gerenciais

Têm por objetivo dar suporte ao nível gerencial através de relatórios, pesquisas nos processos correntes e nos históricos. Orientados a eventos internos, apoiando o planejamento, controle e tomada de decisão. Eles dependem dos Sistemas de Processamento de Transações para a aquisição de dados, o tratamento e a apresentação.

## Sistemas de Apoio Estratégico

A maioria dos executivos das empresas têm pouca ou nenhuma experiência com computadores e são obrigados a utilizar aplicações para apoiar suas tomadas de decisões não rotineiras, que exigem bom senso avaliação e percepção. Criam um ambiente generalizado, de computação e comunicação, em vez de aplicações fixas e capacidades específicas. Projetados para incorporar dados externos como leis e novos concorrentes, também adquirem informações dos Sistemas de Informações da Empresa para apresentar de forma resumida e útil aos executivos.

## Sistemas Transacionais (OLTP)

A informação tem um valor altamente significativo e pode representar grande poder para quem a possui, contém alto valor, pois está integrada com os processos, pessoas e tecnologias, é um ativo intangível como já falamos e é extremamente importante para os negócios.

A empresa passa a ter uma posição estratégica no mercado quando tem a tecnologia como aliada no controle da informação. O fator de sucesso para os negócios é dispor da informação correta, na hora adequada, isto significa tomar uma decisão de forma ágil e eficiente.



## Sistemas Transacionais (OLTP)

A informação é importante tanto na definição quanto na execução de uma estratégia, pois:

- Ajuda a identificar ameaças.
- Ajuda a identificar oportunidades.
- Cria um cenário para uma resposta competitiva.
- Torna-se a matéria-prima se adotar políticas estratégicas eficazes.

## Sistemas Transacionais (OLTP)

Os sistemas transacionais, conhecidos por OLTP (Online Transaction Processing), são a fonte de dados para o Data Warehouse.

A figura ao lado representa o ciclo de vida de um projeto de Data Warehouse.



## Sistemas Transacionais (OLTP)

Vamos estudar a primeira fase: Fonte de Dados, que é composta pelos Sistemas Transacionais (OLTP).

OLTP são aplicações com acesso direto à informação em tempo real. Processam unidades de trabalho, chamadas de transações. Uma única transação pode solicitar uma conta bancária e atualizar o extrato para refletir um depósito.

Exemplos de aplicação: folha de pagamento, controle de vendas, controle de estoque, recursos humanos, etc.

As aplicações comerciais, geralmente, possuem muitos processos similares, por exemplo, geração de um pedido em um sistema de processamento de vendas, uma reserva de assento em um avião ou uma consulta de crédito em um sistema de controle de crédito.

A execução de um desses itens é uma transação. Isso nos leva à definir transação como uma unidade de processamento que corresponde a uma transação comercial e constitui uma entidade lógica dentro de um aplicativo.

## Sistemas Transacionais (OLTP)

Em um sistema de processamento de transações, os usuários finais têm acesso online ao sistema e aos dados da empresa. Em um ambiente de transacional, muitos usuários, repetidamente, executam processos semelhantes e requerem uma resposta rápida para cada operação.

Exemplos de usuários: funcionários que emitem pedidos (balconistas), atendentes de reserva das companhias aéreas ou caixas de banco. Eles compartilham um ambiente de programas e dados em suas respectivas empresas.

## Sistemas Transacionais (OLTP)

Em um típico sistema de processamento de transações, temos:

- Muitos usuários finais executam as mesmas operações, dividindo o mesmo banco de dados e arquivos.
- O sistema pode programar operações com base em atributos de prioridade.
- As operações são invocadas pela entrada de um registro, seu processamento e sua saída.
- As operações são concebidas para uma interface amigável e com tempos de resposta rápidos.
- Acesso imediato aos dados empresariais que tenham sido atualizados refletindo todas as operações anteriores.
- Mudança de dados da empresa deverá refletir imediatamente após cada operação, tal como ela é processada.

## Sistemas Transacionais (OLTP)

A pré-definição de transações permite o seu controle de forma eficiente. Quando executada uma operação envolvendo um ou mais programas, é necessário garantir que no momento de sua execução não aconteça conflito entre programas. O bloqueio impede que transações possam acessar os dados enquanto o ele está sendo atualizado por outra transação

As funções básicas, que devem ser fornecidas por um sistema de processamento de transações, são:

- Processamento online.
- Alta disponibilidade.
- A resposta rápida.
- Baixo custo por transação.
- Acesso e atualização de recursos compartilhados com integridade.

## Sistemas Transacionais (OLTP)

Os sistemas transacionais são base de dados para um projeto de Data Warehouse, então temos que saber qual a diferença entre esses projetos.

O Data Warehouse não substitui estes sistemas, mas os complementa.

## Sistemas Transacionais X Data Warehouse

CARACTERÍSTICAS	SISTEMAS TRANSACIONAIS (OLTP)	DATA WAREHOUSE
Usuários	Operacional e Técnicos	Alta administração
Utilização	Tarefas Cotidianas	Decisões estratégicas
Uso	Operacional	Analítico
Padrão de Uso	Previsível	Difícil de prever
Número de usuários	Grande número de usuários	Poucos usuários
Unidade de trabalho	Inclusão, alteração e exclusão	Inclusão e consulta
Histórico	60 à 90 dias	5 à 10 anos
Integridade	A cada transação	A cada carga
Princípio de funcionamento	Com base em transformações	Com base em análise de dados
Valores de dados	Valores atuais e voláteis	Valores históricos e imutáveis
Detalhamento	Alto	Sumarizado
Organização dos dados	Orientado a aplicações	Orientado ao assunto



## Sistemas Transacionais (OLTP)

Informações típicas de **Sistemas Transacionais**:

Quantas unidades do produto "X" o Sr. Fulano de Tal comprou?

Qual o histórico de pagamento do Sr. Fulano de Tal no último ano?

Quais as cotações realizadas esta semana que se transformaram em vendas esta semana?

## Sistemas Transacionais (OLTP)

Informações que podem ser extraídas de **Sistemas Transacionais e Data Warehouse**:

Qual o volume mensal de vendas em cada estado?

Qual o volume de inadimplência em cada filial da empresa?

Qual o percentual de efetivação de cotações para cada vendedor?

## Sistemas Transacionais (OLTP)

Informações do tipo estratégico que são respondidas por um **Data Warehouse**:

- Qual é nossa posição no mercado nacional, considerando o volume de vendas do produto "X" ?
- Qual o perfil do cliente potencial para o produto "Y"?
- Aumentando as vendas em 20%, teremos mais lucros?
- Quem são meus clientes mais rentáveis?
- Quem são os melhores segmentos de clientes para um novo produto ou serviço?