



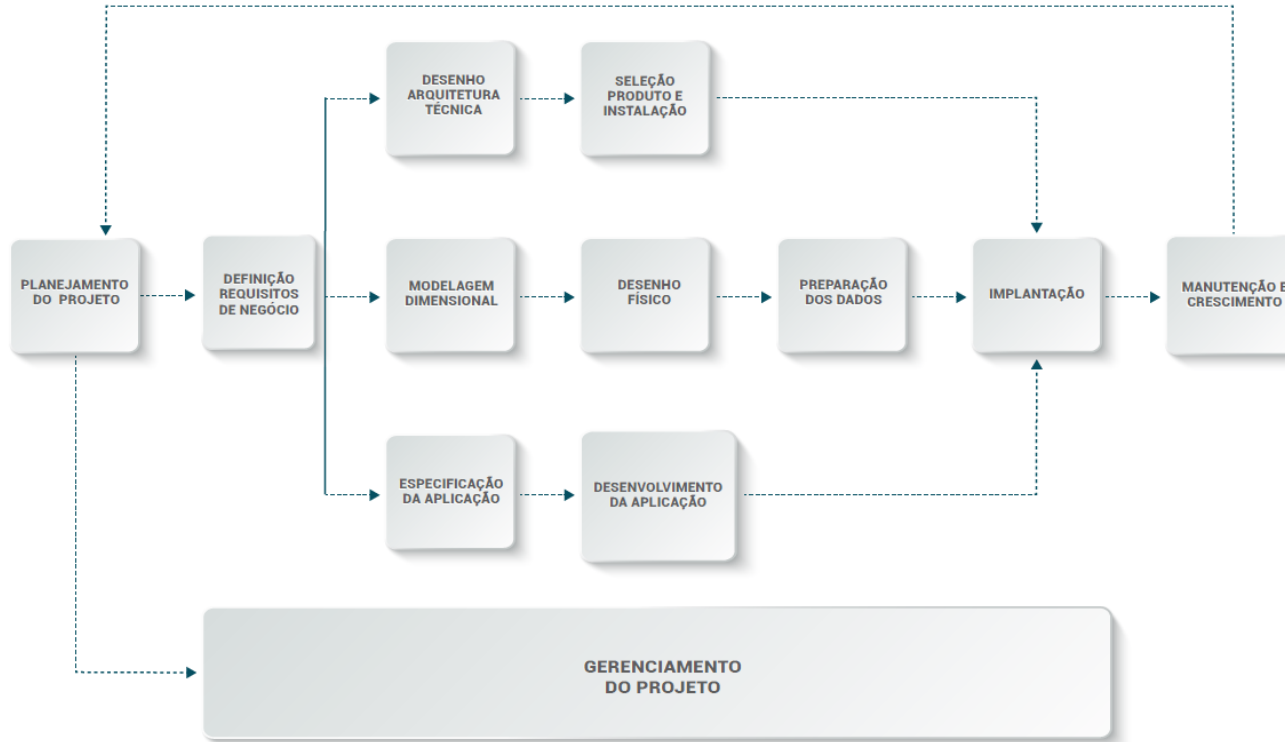
Quem se prepara, não para.

# Business Intelligence

4º período

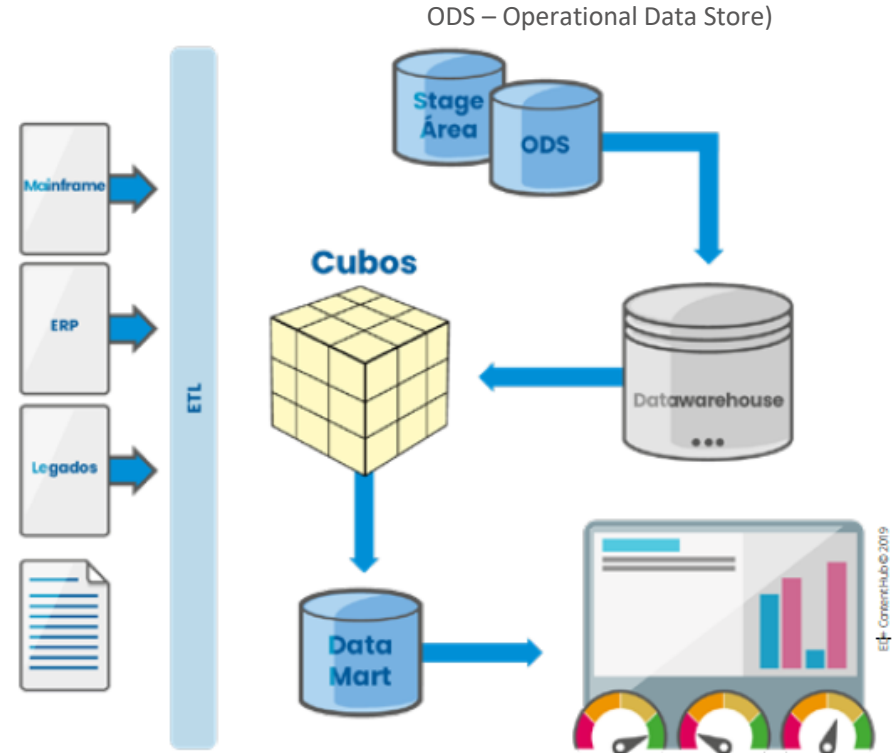
Professora: Michelle Hanne

# Ciclo de Vida Business Intelligence



# Visão de uma Arquitetura de BI

- Uma visão completa de como aplicar os componentes em uma arquitetura de *Business Intelligence*.



# Escopo para Projetos de BI

- **Definição de Requisitos:** Essa fase é de extrema importância e caso você falhe pode colocar tudo a perder. Para que isso não ocorra, você verá técnicas para minimizar falhas, pois somos seres humanos e as falhas são inevitáveis

# Escopo para Projetos de BI

## Definição de Requisitos de Negócio

Os usuários e seus requisitos afetam quase todas as decisões tomadas durante a implementação de um data warehouse. Da sua perspectiva, os requisitos de negócio devem estar no centro do **Ciclo de Vida para Business Intelligence**.



1

### Definição de requisitos

Recomendo que você comece conversando com seus usuários de negócio. Embora isso possa parecer óbvio, muitas empresas utilizam abordagens alternativas que infelizmente, em muitos casos, não vão dar certo. Existem duas técnicas básicas para reunir e levantar requisitos: entrevistas e sessões de brainstorm.

As entrevistas são conduzidas com indivíduos ou grupos muito pequenos.



2

### Entrevista

Deve-se revisar as técnicas para posicioná-lo para o sucesso, eliminando o máximo de incógnitas possíveis durante o processo de coleta de requisitos. Normalmente, estrutura-se uma equipe com os seguintes membros: entrevistador, secretário e corpo técnico, municiando-se sobre o conhecimento da área ou empresa. Selecione os representantes transversais e gere assim questionários específicos para cada perfil/função.



3

### Realizar a entrevista

Revise sempre que possível para os entrevistados algumas regras gerais das entrevistas e respeite as funções e responsabilidades designadas na equipe de entrevista. Compreenda o que o entrevistado está dizendo usando a técnica de parafrasear, padronizando vocabulários que serão usados pelos entrevistados.

Também é fundamental que você resolva qualquer confusão.



4

### Iniciar a entrevista

Agora você está pronto para começar a entrevistar. Toda entrevista começa com apresentações para lembrar ao usuário quem você é e por qual motivo está lá. Dependendo da composição da equipe da entrevista, o gerente do projeto ou o líder do projeto comercial pode comunicar essas informações de alto nível do projeto. Separe os conteúdos das entrevistas conforme o objetivo e papel de cada entrevista: conteúdo para executivos, gerentes, analistas e equipe técnica.

# Escopo para Projetos de BI



5

## Encerre a entrevista

Ao final de cada sessão para levar a entrevista a uma conclusão. O resumo deve começar com uma síntese do que foi discutido durante a sessão.



6

## Revise os resultados da entrevista

Muitas vezes, é útil revisar os relatórios coletados para obter mais informações sobre a dimensionalidade dos negócios e os requisitos analíticos dos usuários. Às vezes você verá dimensões ou atributos nesses relatórios que nunca apareceram em entrevistas.



7

## Preparar e publicar as entregas dos requisitos

A documentação ajuda a equipe a cristalizar e entender melhor o conteúdo da entrevista. Uma coisa é escutar passivamente um entrevistado, mas outra é documentar de forma ativa o que foi dito. Por fim, a documentação permite validar os requisitos achados entre os usuários.



8

## Documento de conclusão dos requisitos

Além das anotações individuais das entrevistas, os dados de todas as entrevistas devem ser consolidados e sintetizados em um documento de constatação de requisitos.



9

## Concorde nas próximas etapas após a coleta de requisitos

Você tem uma noção do que os usuários desejam e do que será necessário para atender a esses requisitos. Primeiro os usuários precisam confirmar se as descobertas dos requisitos sintetizados são precisas. Segundo, a menos que seus usuários sejam significativamente diferentes da maioria com os quais você trabalha, eles inevitavelmente desejam mais do que podem ser entregues em uma única fase da implementação do seu armazém de dados.

# Sistemas Transacionais vs Sistemas Analíticos



# Sistemas Transacionais

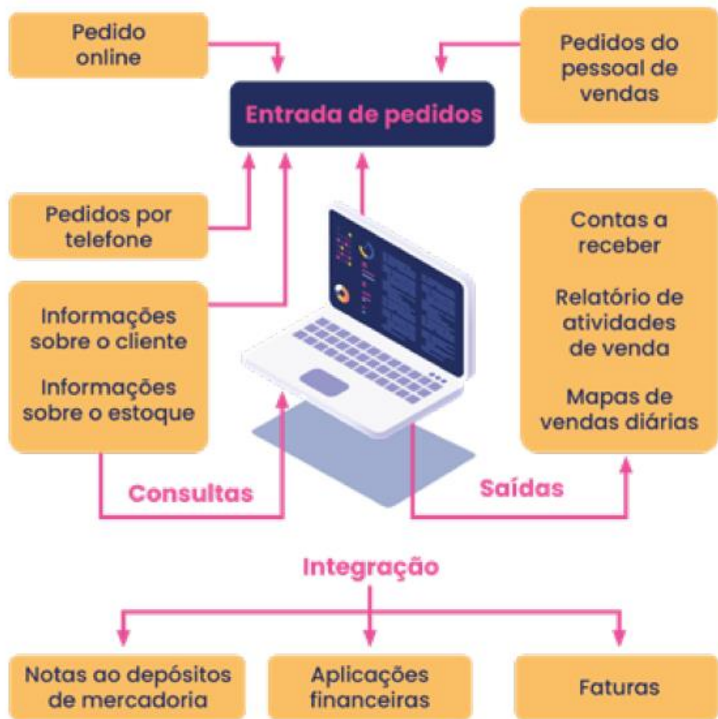
Os Sistemas de Informações Transacionais dão suporte às atividades diárias da empresa.

O objetivo principal do SIT é o fornecimento de todas as informações legais ou organizacionais referentes à empresa para manter eficientemente os seus negócios.

As principais vantagens são: a precisão e a confiabilidade obtidas, a redução no custo e o tempo de obtenção das informações.

# Sistemas Transacionais

## Processamento de pedidos



Os Sistemas de Informações Transacionais são **estáticos por natureza**, eles mudam apenas em resposta a uma mudança intencional nas políticas ou processos de negócios, ou por razões técnicas, como manutenção do sistema ou ajuste de desempenho.

# OLTP (*Processamento de Transações Online*)

Os sistemas OLTP registram as interações de negócios conforme elas ocorrem na operação diária da organização e dão suporte à consulta desses dados para criar inferências.

**O OLTP requer suporte para transações em rede**, por isso, os softwares OLTP utilizam processamento cliente/servidor e softwares intermediários que permitem às transações rodarem em diferentes plataformas de computadores numa rede.

Em grandes aplicações, a eficiência do OLTP vai depender de um **sofisticado software de gerenciamento de transações** (como o CICS) e/ou de otimizações táticas de base de dados de um grande número concorrente de updates em uma base de dados orientado ao OLTP.

# OLTP (*Processamento de Transações Online*)

## Exemplos de transações OLTP:

- Banco online;
- Compra de um livro online;
- Reserva de passagem aérea;
- Envio de uma mensagem de texto;
- Entrada de pedidos;
- Operadores de telemarketing inserindo resultados de pesquisas por telefone;
- Equipe da central de atendimento visualizando e atualizando os detalhes dos clientes.

## ERP

- ERP é um Sistema de Informação Transacional que interliga todos os dados e os processos de uma organização em um único sistema.
- A interligação pode ser vista sob a **perspectiva funcional** (sistemas de finanças, contabilidade, Recursos Humanos, fabricação, Marketing, vendas e compras) e sob a **perspectiva sistêmica (Sistema de Processamento de Transações)**.
- O ERP é uma plataforma de software desenvolvida para **interligar diversos departamentos de uma empresa**, possibilitando a automação e o armazenamento de todas as informações do negócio.

# Sistemas Analíticos

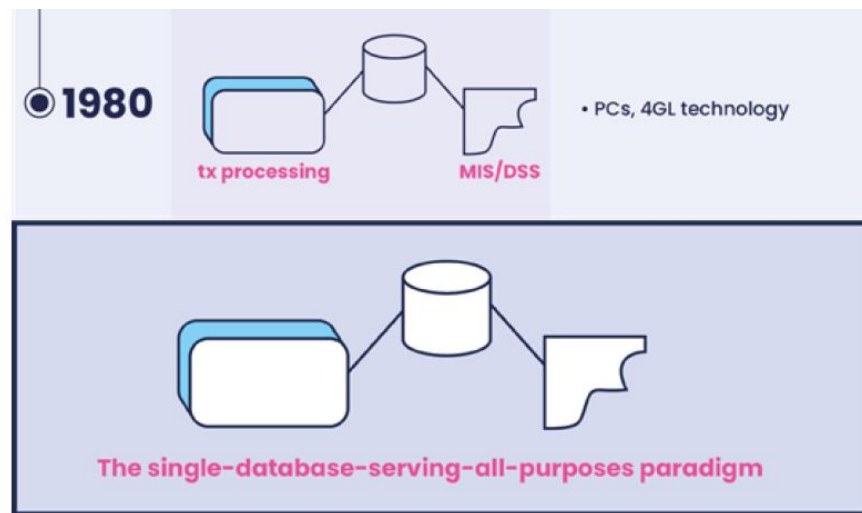
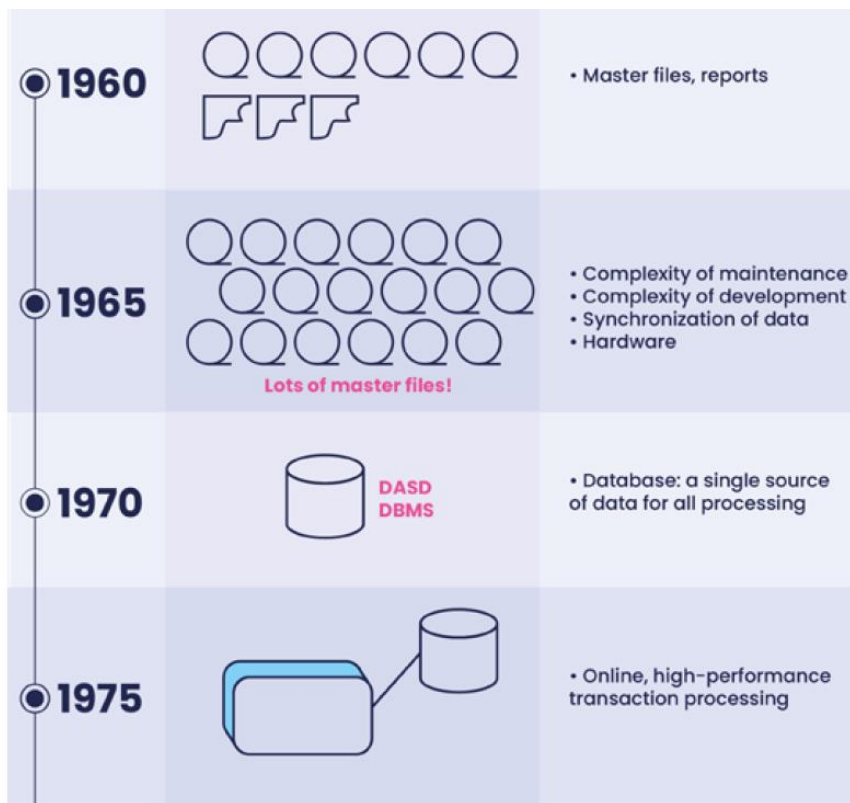


Figura 2 - Linha do tempo para construção de data warehouse de Business Intelligence.  
Fonte: adaptado de W. H. Inmon (2005).

- Em meados de 1960, o crescimento dos arquivos mestres e da fita magnética explodiu. E com esse crescimento vieram grandes quantidades de dados redundantes. A proliferação de arquivos mestres e dados redundantes apresentou alguns problemas:
  - A necessidade de sincronizar dados após a atualização;
  - A complexidade de manter programas;
  - A complexidade do desenvolvimento de novos programas;
  - A necessidade de grandes quantidades de hardware para suportar todos os arquivos principais.

SISTEMA DE SUPORTE À DECISÃO (DSS)

- A **década de 1970 assistiu ao advento** do armazenamento em disco ou do DASD (Direct Access Storage Device). O armazenamento em disco era fundamentalmente diferente do armazenamento em fita magnética, pois os dados podiam ser acessados diretamente em um DASD.
- Com o DASD surgiu um novo tipo de software de sistema conhecido como **Sistema de Gerenciamento de Banco de Dados (DBMS)**. O objetivo do DBMS era facilitar para o programador armazenar e acessar dados em um DASD.
- Já na **década de 1980**, você pode acompanhar que as mais novas tecnologias, como PCs e linguagens de quarta geração (4GLs), começaram a surgir. O usuário final começou a assumir uma função anteriormente incompreensível – controlando diretamente dados e sistemas – e preservada para o processador de dados profissional. Sistema de Informações de Gerenciamento (MIS)



# OLAP (*Online Analytical Processing*)

OLAP é um conceito de interface com o usuário, que proporciona a capacidade de ter ideias sobre os dados, permitindo analisá-los profundamente em diversos ângulos. As funções básicas do OLAP são:

- Visualização multidimensional dos dados;
- Exploração;
- Rotação.

Os dados a serem analisados em um ambiente OLAP são organizados de maneira a permitir visibilidade em qualquer uma das dimensões.

# OLAP (*Online Analytical Processing*)

Geralmente, isso é **descrito como um cubo**, embora a organização pretenda permitir que o analista fixe algum conjunto de dimensões e depois veja agregados associados às outras hierarquias dimensionais. **Exemplo considere uma tabela de fatos de vendas que registra todas as transações de vendas, incluindo data, hora, local, cliente, item, quantidade, preço por produto, balconista, promoção de vendas e total de vendas.**

Você pode configurar um cubo OLAP com as seguintes dimensões:

- Cliente;
- Local de vendas;
- Produto;
- Tempo;
- Balconista;
- Promoção de vendas.

# OLAP (*Online Analytical Processing*)



Camadas de tomadas de decisão. **Fonte:** adaptado de W. H. Inmon (2005).

EIS – Sistema de Informação Executiva

MIS - Sistema de Informações de Gerenciamento

DSS - Sistema de Suporte a Decisões

## OLAP

(Online Analytical Processing)

O OLAP (Online Analytical Processing) é a capacidade para manipular e analisar um grande volume de dados sob múltiplas perspectivas. As aplicações OLAP são usadas pelos gestores em qualquer nível da organização para lhes permitir análises comparativas que facilitem a sua tomada de decisões diárias.



### ROLAP (Relational Online Analytical Processing)

Utiliza a tecnologia de banco de dados relacionais para armazená-los e é usado para grandes volumes de dados. No ROLAP, a visão multidimensional estática dos dados é criada.



### MOLAP (Multidimensional Online Analytical Processing)

É um conjunto de ferramentas que dispara suas requisições diretamente ao servidor de banco de dados multidimensional. O MOLAP é usado para volumes de dados limitados, os quais são armazenados em uma matriz multidimensional.



### HOLAP (Hybrid Online Analytical Processing)

É um conjunto de ferramentas híbridas. É a combinação entre ROLAP e MOLAP, pegando o melhor de ambas as categorias: a escalabilidade do ROLAP e o alto desempenho do MOLAP.



### WOLAP (Web Online Analytical Processing)

É um conjunto de ferramentas que dispara uma consulta via navegador web para o servidor, o qual retorna enviando o cubo processado de volta, para que possa ser analisado pelo usuário.



### DOLAP (Desktop Online Analytical Processing)

É um conjunto de ferramentas que dispara uma consulta de uma estação-cliente para o servidor, o qual retorna enviando o macrocubo de volta, e que foi projetado para usuários departamentais simples e de baixo custo. Os dados são armazenados em cubos na área de trabalho; é como ter sua própria planilha. Já que os dados são locais, os usuários finais não precisam se preocupar com impactos no desempenho contra o servidor.

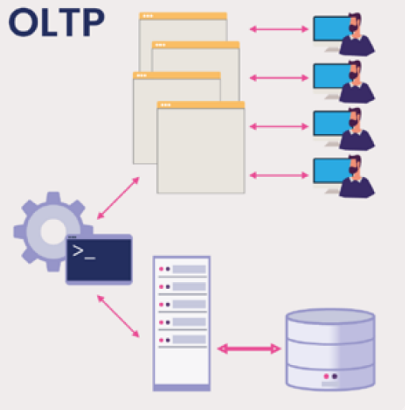
# Transacionais vs Analíticos

O sistema **OLTP** foca no nível operacional da organização. Ele colhe informações operacionais, edita estes dados para manter o fluxo do sistema, armazena essas informações em sistemas convencionais de bancos de dados e auxilia na confecção de relatórios.

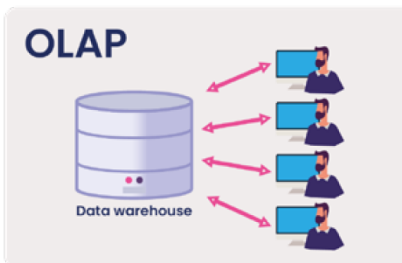
Já o **OLAP** auxilia empresas a organizarem seus bancos de dados e dão suporte para que elas façam uma análise profunda dessas informações. A sua utilização faz com que as organizações possam acessar, visualizar e analisar esses dados para extrair insights.

# Transacionais vs Analíticos

## OLTP



## OLAP



EDI Content Hub © 2019

	OLAP	OLTP
Foco	Foco no nível estratégico da organização. Visa a análise empresarial e a tomada de decisão.	Foco no nível operacional da organização. Visa a execução operacional do negócio.
Performance	Otimização para a leitura, como também para a geração de análises e relatórios gerenciais.	Alta velocidade na manipulação de dados operacionais, porém ineficiente para a geração de análises gerenciais.
Estrutura dos dados	Os dados estão estruturados na modelagem dimensional. Eles normalmente possuem um alto nível de sumarização.	Os dados são normalmente estruturados em um modelo relacional normalizado, otimizado para a utilização transacional. Eles possuem alto nível de detalhes.
Armazenamento	O armazenamento é feito em estruturas de Data Warehouse com otimização no desempenho em grandes volumes de dados.	O armazenamento é feito em sistemas convencionais de banco de dados por meio dos Sistemas de Informações da organização.
Abrangência	É utilizado pelos gestores e analistas para a tomada de decisão.	É utilizado por técnicos e analistas, como também engloba vários usuários da organização.
Frequência de atualização	A atualização das informações é feita no processo de carga dos dados. A frequência é baixa, podendo ser diária, semanal, mensal ou anual (ou critério específico).	A atualização dos dados é feita no momento da transação. A frequência das atualizações é muito alta.
Volatilidade	Dados históricos e não voláteis. Os dados não sofrem alterações, salvo necessidades específicas (por motivos de erros ou inconsistências de informações).	Dados voláteis, passíveis de modificação e exclusão.
Tipos de permissões nos dados	Apenas a inserção e a leitura são permitidas, sendo que para o usuário está apenas disponível a leitura.	A leitura, a inserção, a modificação e a exclusão dos dados podem ser feitas.

# Modelagem de Dados

# Data Warehouse

Data Warehouse (DW) é um conjunto de dados organizados de forma analítica para oferecer suporte à tomada de decisões.

Definido de várias maneiras diferentes, mas não de uma forma rigorosa.

- Uma base dados de suporte a decisão que é mantida separadamente da base operacional da organização.
- Suporta processamento de informação fornecendo uma plataforma sólida para análise de dados históricos, consolidados.

Data Warehouse é um subconjunto de um Data Warehouse. Normalmente é menor e se concentra em um assunto ou departamento específico, porém pode unir vários bancos de dados de uma empresa.



# Data Warehouse

Organizado em torno de temas importantes, tais como: cliente, produto, vendas.

Focado na modelação e análise de dados para quem toma decisões, em vez de operações diárias e processamento de transações.

Fornece uma visão simples e concisa sobre questões de um tema particular através da exclusão de dados que não são importantes no suporte ao processo de decisão.

- O horizonte de tempo para um Data Warehouse é significativamente maior do que o de SGBDs:
  - Base de dados operacional: informação atual.
  - Dados no Data Warehouse: informação em uma perspectiva histórica (exemplo: últimos 5 -10 anos)
- Cada estrutura chave no DW contém um elemento de tempo, explicitamente ou implicitamente.

# Por que separar um DW de um SGBD

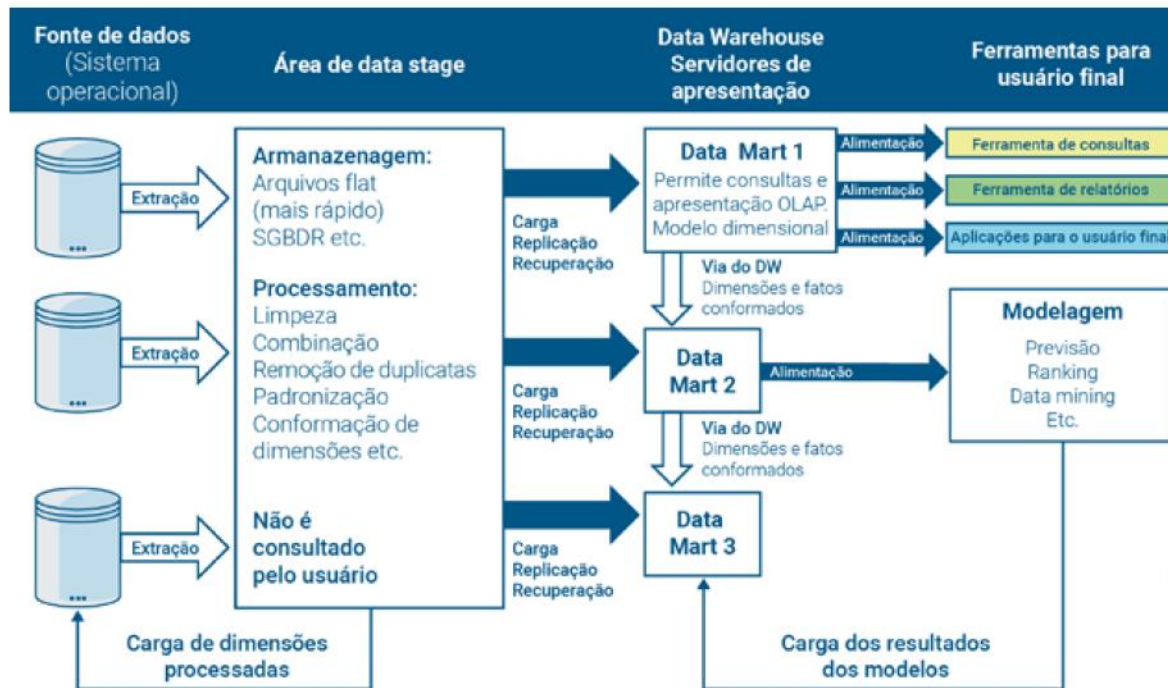
- Alta performance para ambos:
  - **SGBD**: otimizado para OLTP – métodos de acesso, indexação, controle de concorrência, recuperação, etc.
  - **DW**: otimizado para OLAP – consultas complexas, visões multi-dimensionais, consolidação, etc.

# Por que separar um DW de um SGBD

Essas preocupações são tão universais que direcionam os requisitos básicos para o data warehouse. Você deve transformar essas reclamações de gerenciamento de negócios em requisitos de data warehouse:

- O data warehouse deve apresentar as informações da organização de maneira consistente.
- O data warehouse deve ser adaptável e resiliente à mudança.
- O data warehouse deve ser um bastião seguro que proteja nossos ativos de informações.
- O data warehouse deve servir como base para uma melhor tomada de decisão.
- A comunidade comercial deve aceitar o data warehouse para que seja considerado bem-sucedido.

# Componentes de um DW



ED+ Content Hub © 2019

**Figura 4 - Componentes do data warehouse.**  
Fonte: adaptado de Ralph Kimball e Margy Ross (2002).

# Componentes de um DW

**Sistemas operacionais de origem:** esses são os sistemas de *ERP* que capturam as transações dos negócios. Os sistemas de origem devem ser vistos como fora do data warehouse, porque presumivelmente se tem pouco ou nenhum controle sobre o conteúdo e a forma dos dados nesses sistemas operacionais legados. As principais prioridades dos sistemas de origem são desempenho e disponibilidade do processamento.

**Área de data stage:** é uma área de armazenamento e um conjunto de processos comumente referido como **ETL** (Extração-Transformação-Carga). A extração é a primeira etapa do processo de inserção de dados no data warehouse. Extrair significa ler e entender os dados fornecidos e copiar os dados necessários para o data warehouse na área de data stage para manipulação adicional.

**Data warehouse servidores de apresentação:** essa área é onde os dados são organizados, armazenados e disponibilizados para consultas diretas por usuários, relatórios e outros aplicativos analíticos.

**Ferramentas para o usuário final:** usa-se o termo ferramenta para se referir à variedade de capacidades que podem ser fornecidas aos usuários de negócios para alavancar a tomada de decisões analíticas: relatórios, dashboards, alertas, etc.

# Modelagem Multidimensional

Essa técnica é utilizada em projetos de BI, aplicados nos dados relacionais. A modelagem multidimensional (**também conhecida como modelagem dimensional**) é baseada nas estruturas de **Fatos e Dimensões**. Esse modelo de trabalho é ideal para estruturação de dados em um **Data Warehouse (DW)**.

A dimensão é o que dá personalidade e qualidade aos “Fatos” ocorridos, é a dimensão que nos permite visualizar as informações por diversos aspectos.

Assim é possível estruturar os dados em cubos (junção entre dimensões e fatos) por assuntos, de forma a entregar mais resultados para o usuário que vai consumir essas informações.

# Modelagem Multidimensional

Essa técnica é utilizada em projetos de BI, aplicados nos dados relacionais. A modelagem multidimensional (**também conhecida como modelagem dimensional**) é baseada nas estruturas de **Fatos e Dimensões**. Esse modelo de trabalho é ideal para estruturação de dados em um **Data Warehouse (DW)**.

A dimensão é o que dá personalidade e qualidade aos “Fatos” ocorridos, é a dimensão que nos permite visualizar as informações por diversos aspectos.

Assim é possível estruturar os dados em cubos (junção entre dimensões e fatos) por assuntos, de forma a entregar mais resultados para o usuário que vai consumir essas informações.

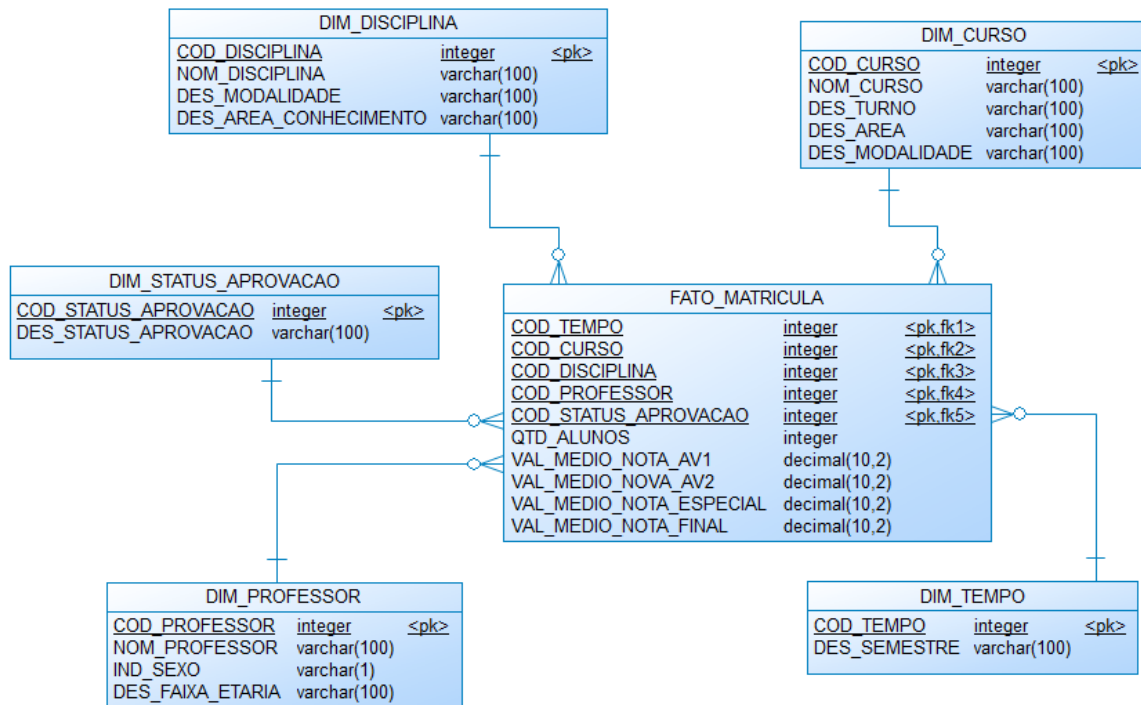


# Modelagem Multidimensional

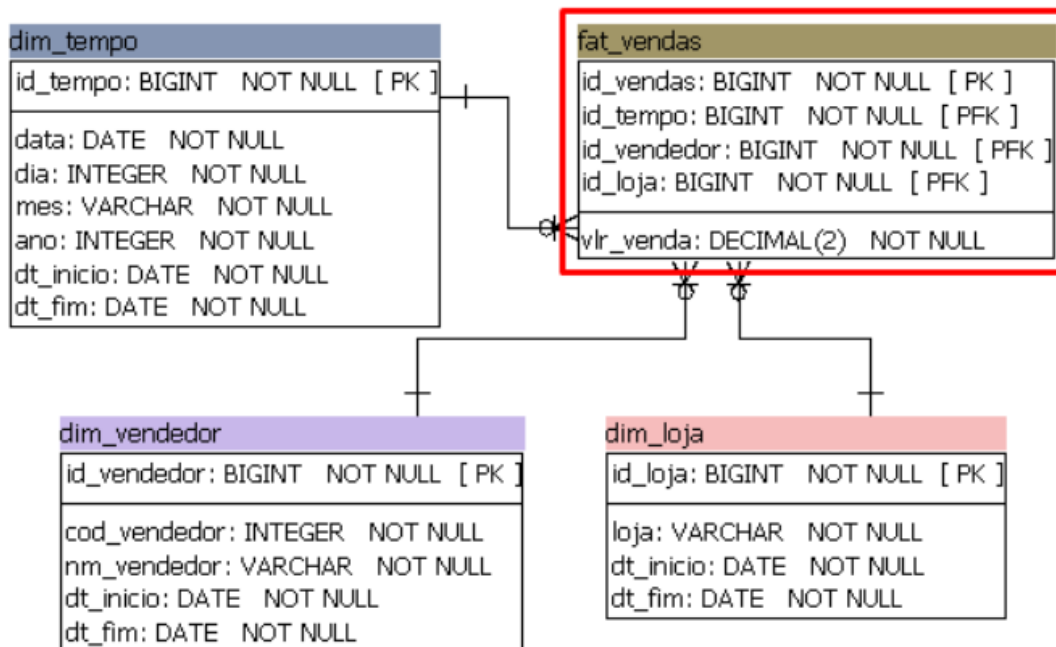
**Tabela Fato:** São aquelas que nos trazem fatos ou eventos que ocorreram. **Exemplos desse tipo de tabela são vendas, saldos, ordens e outros valores quantitativos.** Como característica, nossa **Tabela** normalmente **contém milhões de linhas**, e está continuamente incluindo registros. **Também é muito comum ter uma coluna com informações de data.** Outra informação muito importante é que nossa **Tabela Fato** nos trará a **chave conhecida como *Primary Key*, que será a chave de relação com o nosso outro tipo de tabela.** Esta tabela é o coração do relatório.

**Tabela Dimensão:** São tabelas que vão nos trazer pontos nos quais os fatos serão analisados. Na nossa **dTabela**, **teremos descrições dos eventos que foram trazidos em nossa Tabela Fato**, e não seria errado pensar em Tabela Dimensão como cadastros. Também temos nossa coluna ***Primary Key*, chave para fazer o relacionamento entre as tabelas Dimensão e Fato**, contendo informações com registros únicos como o ID. Como exemplos de dimensão temos: datas, produtos, países, clientes etc.

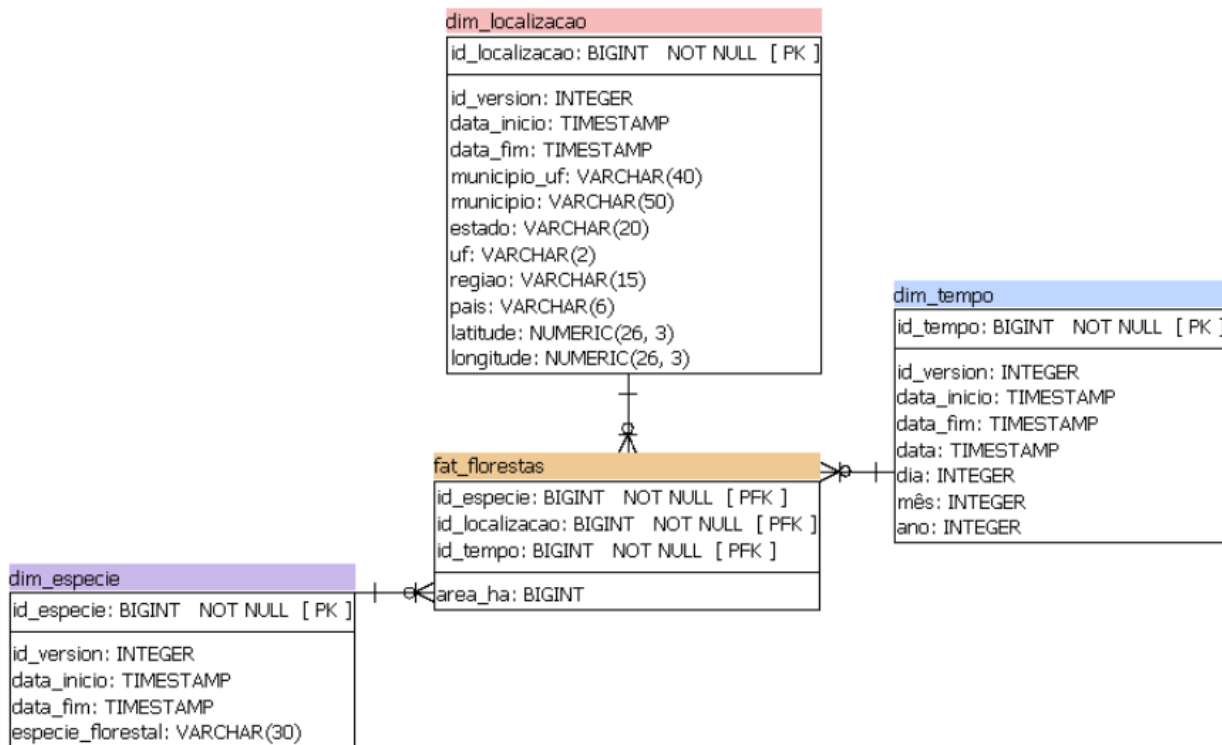
# Modelagem Multidimensional



# Modelagem Multidimensional



# Modelagem Multidimensional



# Tutorial Florestas Plantadas

**Pergunta de Negócio:** investigar quais tipos de espécies foram plantadas entre os anos de 2014 e 2016 e em quais locais do país.

Será que o país já se beneficiou com o investimento de plantação?  
Qual a relação de área de coberta vs desmatamento? Será que estados que receberam maiores plantações desmatam mais?

Baixar o arquivo **rf\_florestasplantadas\_ibge** no site do IBGE:  
<https://dados.agricultura.gov.br/ne/dataset/snif/resource/65cce855-685e-4cfa-b3a4-2d192fc83b4e>

# Referências

- GONÇALVES, Glauber Rogério Barbieri. Sistemas de Informação. Porto Alegre: SAGAH, 2017. ISBN digital 9788595022270.
- INMON, W. H. Building the Data Warehouse. 4. ed. Indianápolis: Wiley, 2005. ISBN 0-7645-9944-5.
- MORAIS, I. S. et al. Introdução a Big Data e Internet das Coisas. Porto Alegre: SAGAH, 2018. ISBN digital 9788595027640.
- STAIR, Ralph M.; REYNOLDS, George W. Princípios de Sistemas de Informação. 11. ed. São Paulo: Cengage Learning, 2016. ISBN digital 9788522124107.
- TURBAN, Efraim et al. Business Intelligence: um enfoque gerencial para a inteligência do negócio. Porto Alegre: Bookman, 2009. ISBN digital 9788577804252.