

A decorative graphic on the left side of the slide consisting of a grid of squares in various shades of blue and purple, arranged in a stepped pattern.

Fundamentos de *Business Intelligence* (1º Encontro)

Prof. MSc. Fernando Siqueira

Fernando Siqueira

Formação Acadêmica



Especialista em Gerenciamento de Projetos



Mestre em Informática Aplicada



Bacharel em Informática



Téc. em Informática Industrial

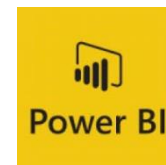
Fernando Siqueira

Experiência Profissional



Fernando Siqueira

Experiência Profissional



Apresentação da Turma

- Experiência profissional
- Expectativa do Módulo



Orientações



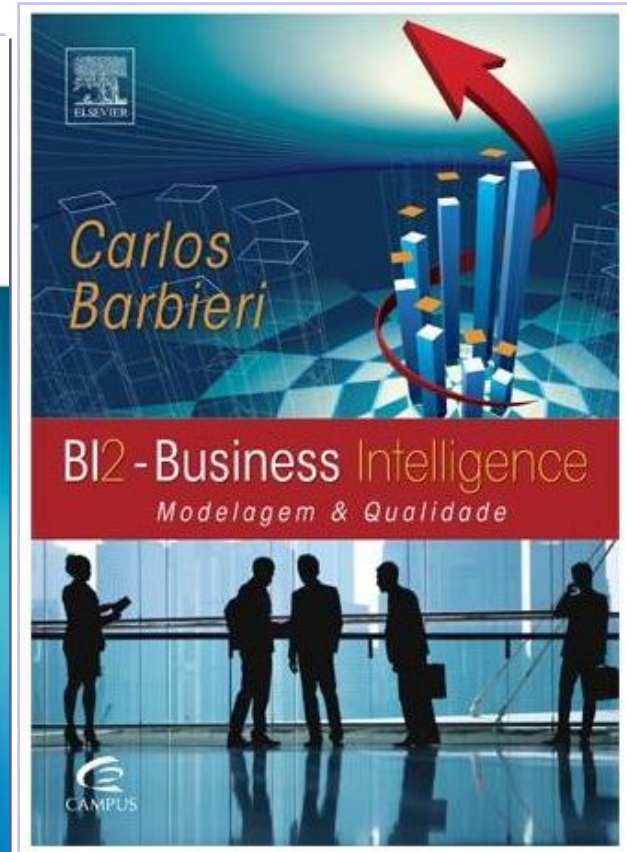
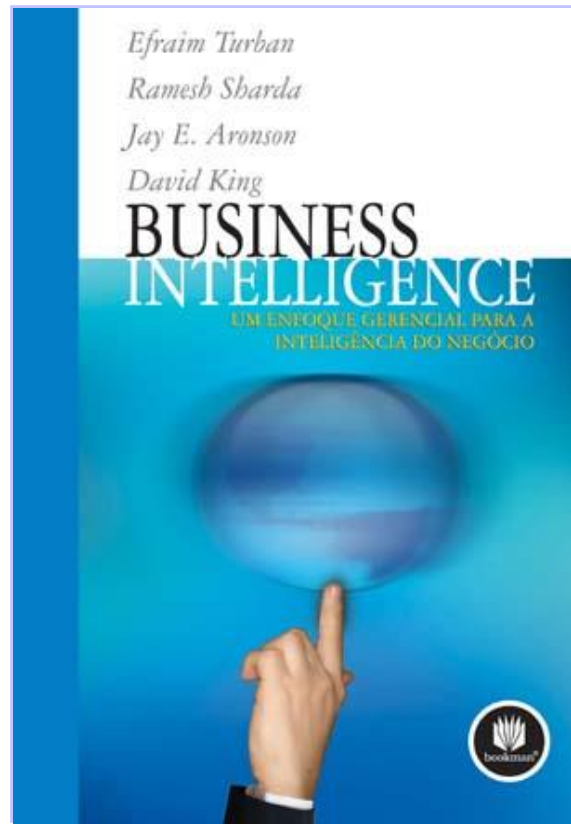
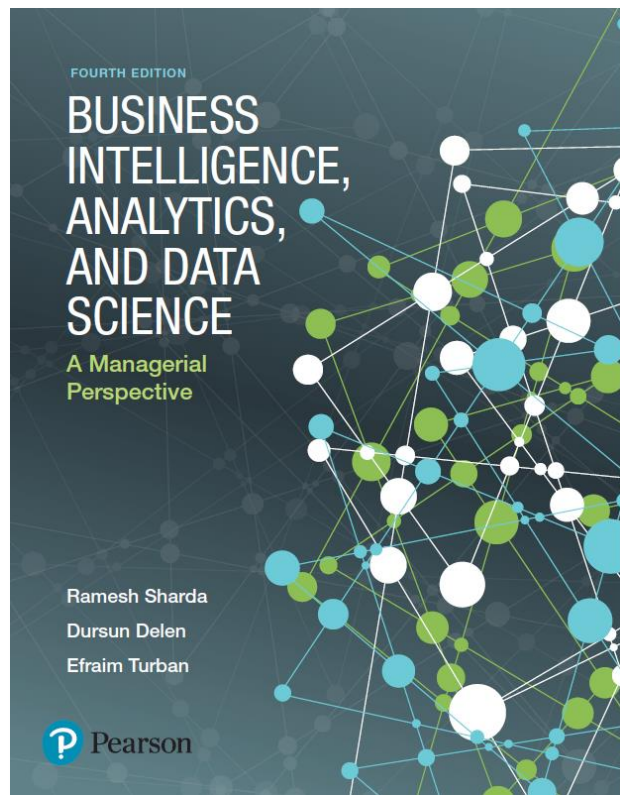


Objetivos do Módulo

- Qualificar os profissionais para construção de soluções de *Business Intelligence*, a partir de sua capacitação em ferramentas, arquiteturas, *Data Warehouse*, *Data Mart*, ODS, componentes de uma plataforma de BI, baseados em um processo de obtenção, análise e distribuição de informações. Destacando a importância do BI e sua relação com administração da empresa.
- 03 encontros nos dias: 30/09, 21/10 e 28/10.



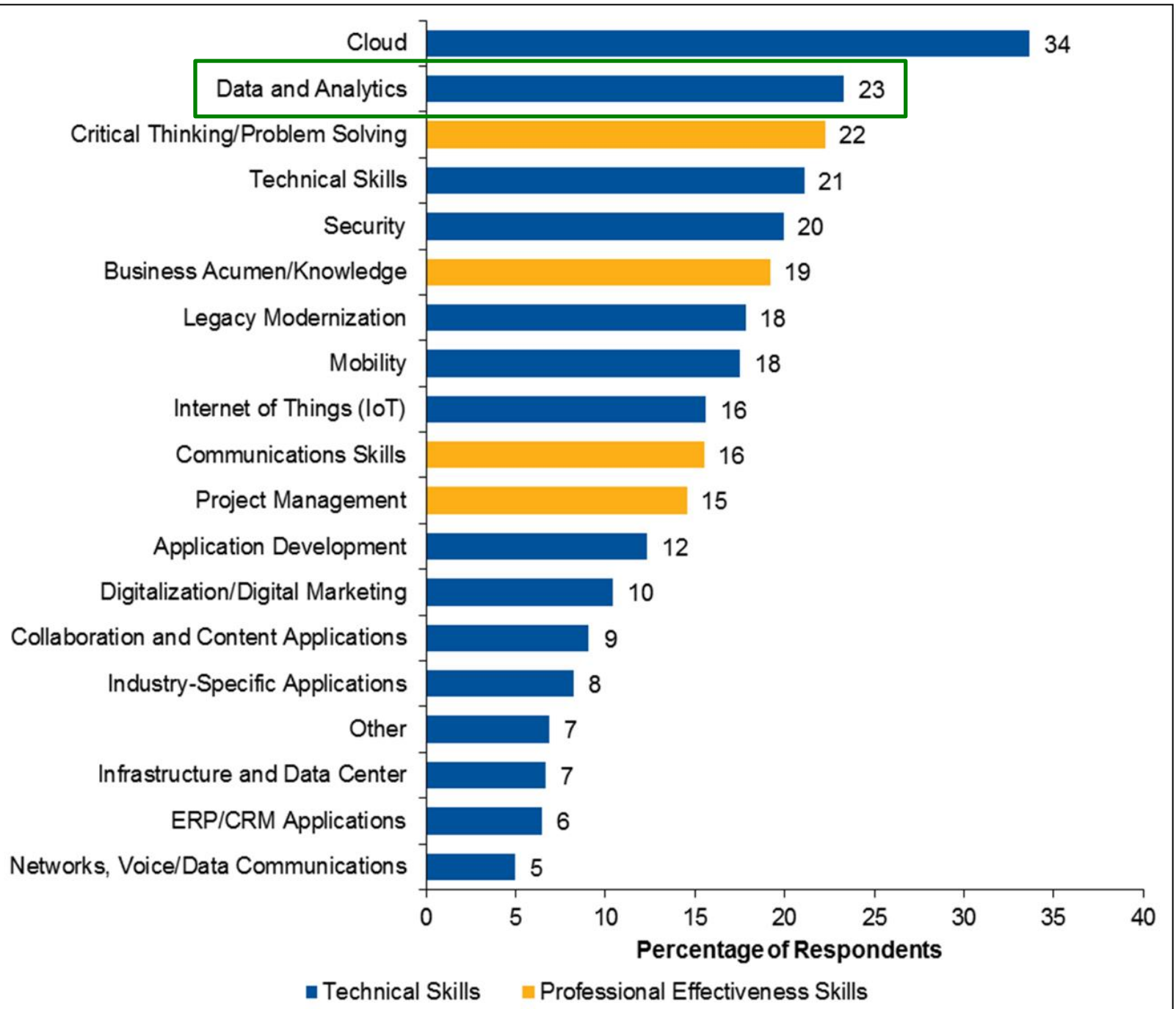
Referência Bibliográfica



Sumário

- Visão Geral
- *Data Warehouse*
- *Data Mart*
- *Business Intelligence*
- Ferramentas de Visualização de Dados
- *Self-service BI*
- *Logical Data Warehouse*





Visão Geral



- *In 2017, analytics will go viral within and outside the enterprise. Technical professionals will need to holistically manage their data and analytics architecture from end to end and leverage cloud wherever appropriate to meet the requirement for "analytics everywhere."*

(John Hagerty , 2016)

Conceitos



- Processos inteligentes de coleta, organização, análise, compartilhamento e monitoração de dados contidos em **Data Warehouse** (DW) ou **Data Mart** (DM), gerando informações para o suporte à tomada de decisões no ambiente empresarial.
- **BI não é um produto** novo ou uma tecnologia, mas sim **um conjunto de conceitos, metodologias, ferramentas, banco de dados** que, fazendo uso de dados extraídos de uma organização, apoia a tomada de decisões.

Conceitos



- **Business Intelligence** é um conceito muito atual que vai além da gestão empresarial. Entre outras coisas, envolve a utilização de **produtos** e **soluções** com tecnologia analíticas de ponta que permitem transformar dados armazenados em bases de dados em informações que auxiliam os diversos níveis de uma empresa na tomada de decisões.
- **Inteligência** é o produto da transformação de dados em informação, após ser analisada ou inserida em um determinado ambiente. Esta informação transformada, aplicada a um determinado processo de decisão, gera **vantagem competitiva** para a organização.

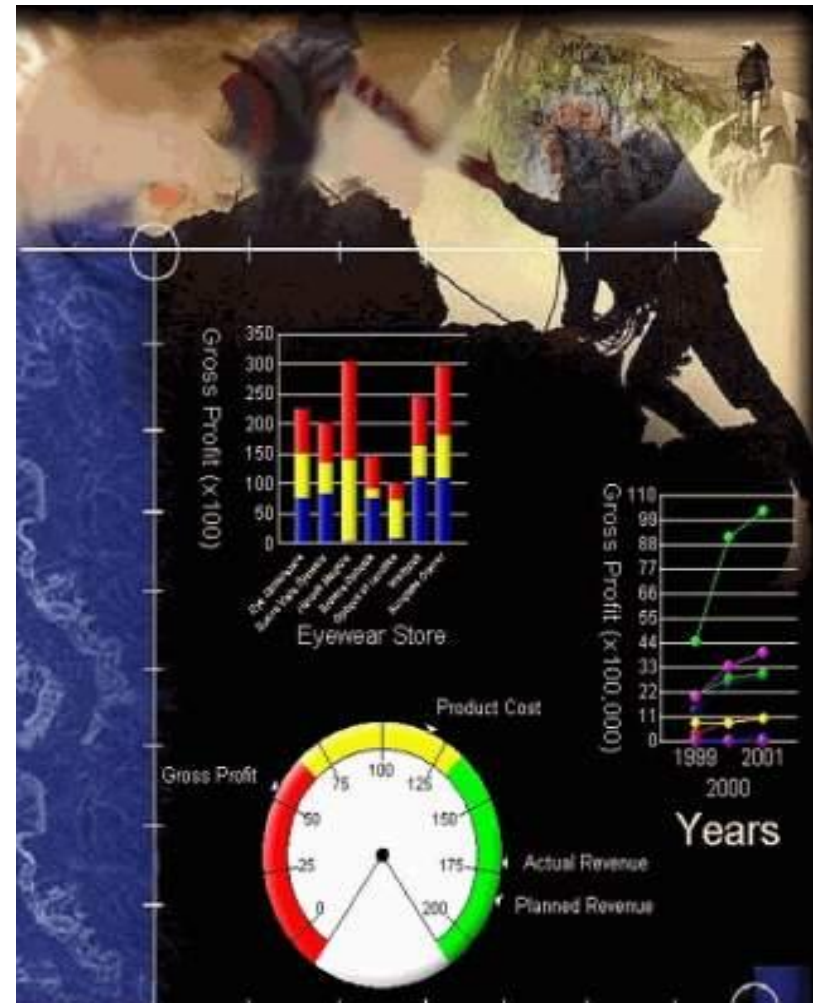
Objetivos

- Disponibilizar informações da maneira e formato correto e no tempo certo para que a empresa possa tomar decisões melhores e mais rápidas.
- Permitir a realização de análises e projeções, de forma a melhorar os processos relacionados às tomadas de decisão e ao ganho de vantagens competitivas.



Benefícios e Vantagens

- As ferramentas de *Business Intelligence* possibilitam a busca e interpretação de informações armazenadas na corporação, garantindo maior exatidão nas tomadas de decisão;
- Permite à Gerência estabelecer uma aproximação integrada e colaborativa para capturar, criar, organizar e usar todos os ativos de informação de uma empresa;
- Antecipação às mudanças bruscas no mercado;
- Antecipação às ações sobre os competidores;
- Conhecimento sobre o negócio;
- Aprendizado através do sucesso e falhas dos concorrentes;
- Visão clara sobre novos negócios;
- Auxílio na implementação de novas ferramentas gerenciais;
- O **Business Intelligence**, como interface, interfere, transforma e torna verdadeira todas estas informações e as transforma em conhecimento estratégico.



Benefícios e Vantagens

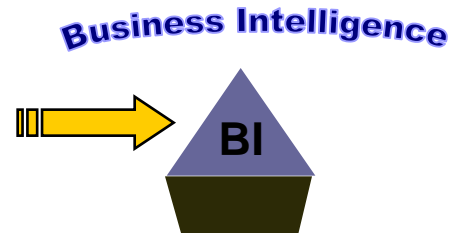
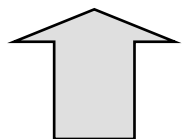
- Antecipa mudanças de mercado
- Antecipa ações dos competidores
- Descobre novo ou potenciais clientes
- Entra em novos negócios
- Possibilita a revisão de suas práticas de negócios
- Aumenta seu grau de competitividade



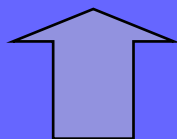
Business Intelligence

Estrutura

Dos Dados Operacionais para
Informações gerenciais



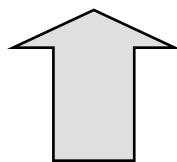
Dados organizados para os modelos de
negocio



DATA WAREHOUSE



Integração e Validação dos dados



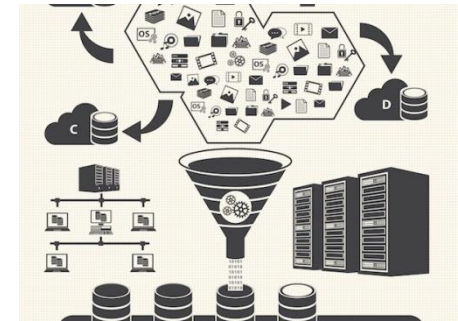
Banco de Dados
Operacional - ERP

Dados Operacionais



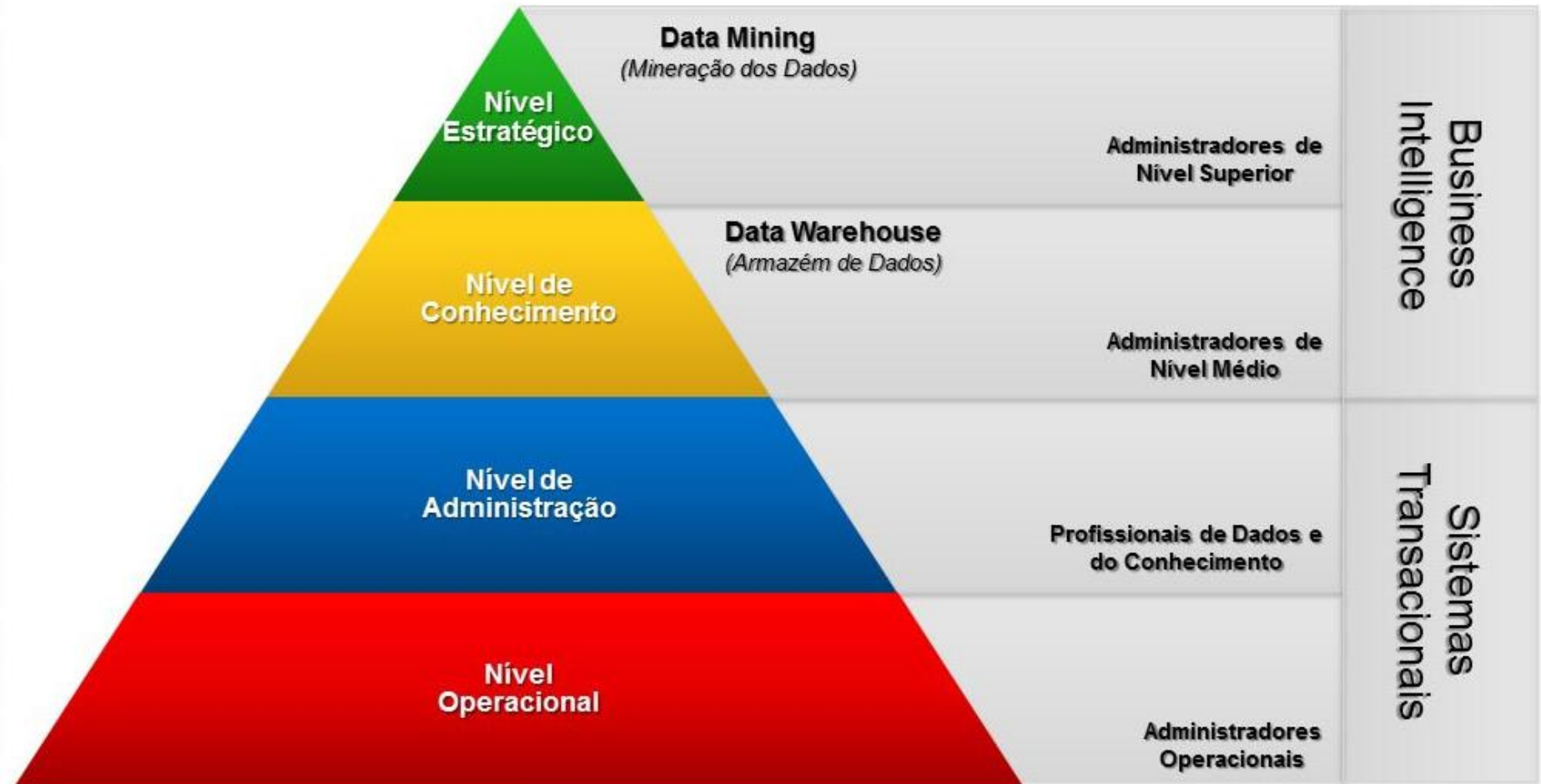
Dados Externos
Fontes diversas

Uso dos Dados

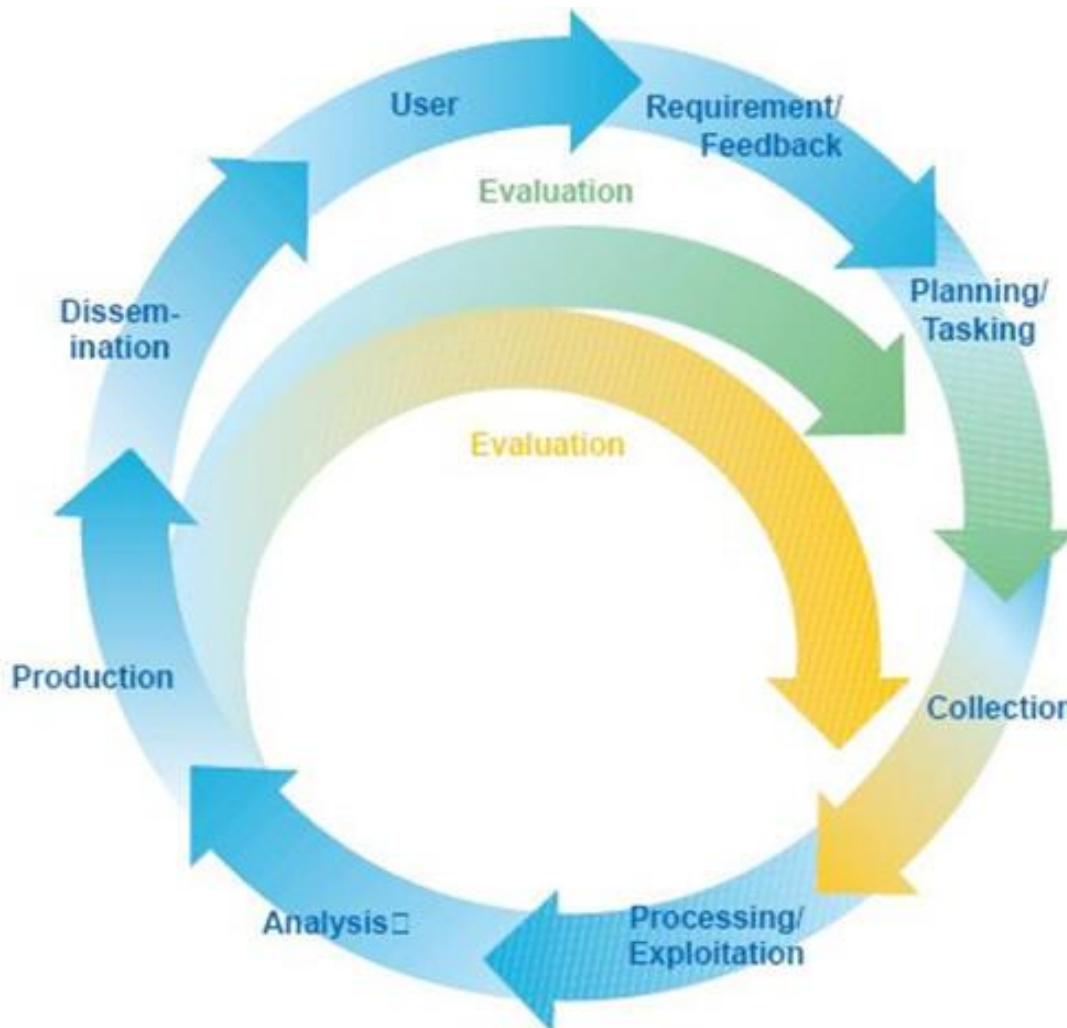


- Os dados são a matéria-prima para qualquer decisão. Vêm de dentro e fora da empresa.
- Existem em toda parte: em repouso, em movimento, no local e na nuvem.
- A análise de dados está assumindo um papel mais ativo e dinâmico ao impulsionar as atividades de toda a organização.

Níveis de Usos dos Dados

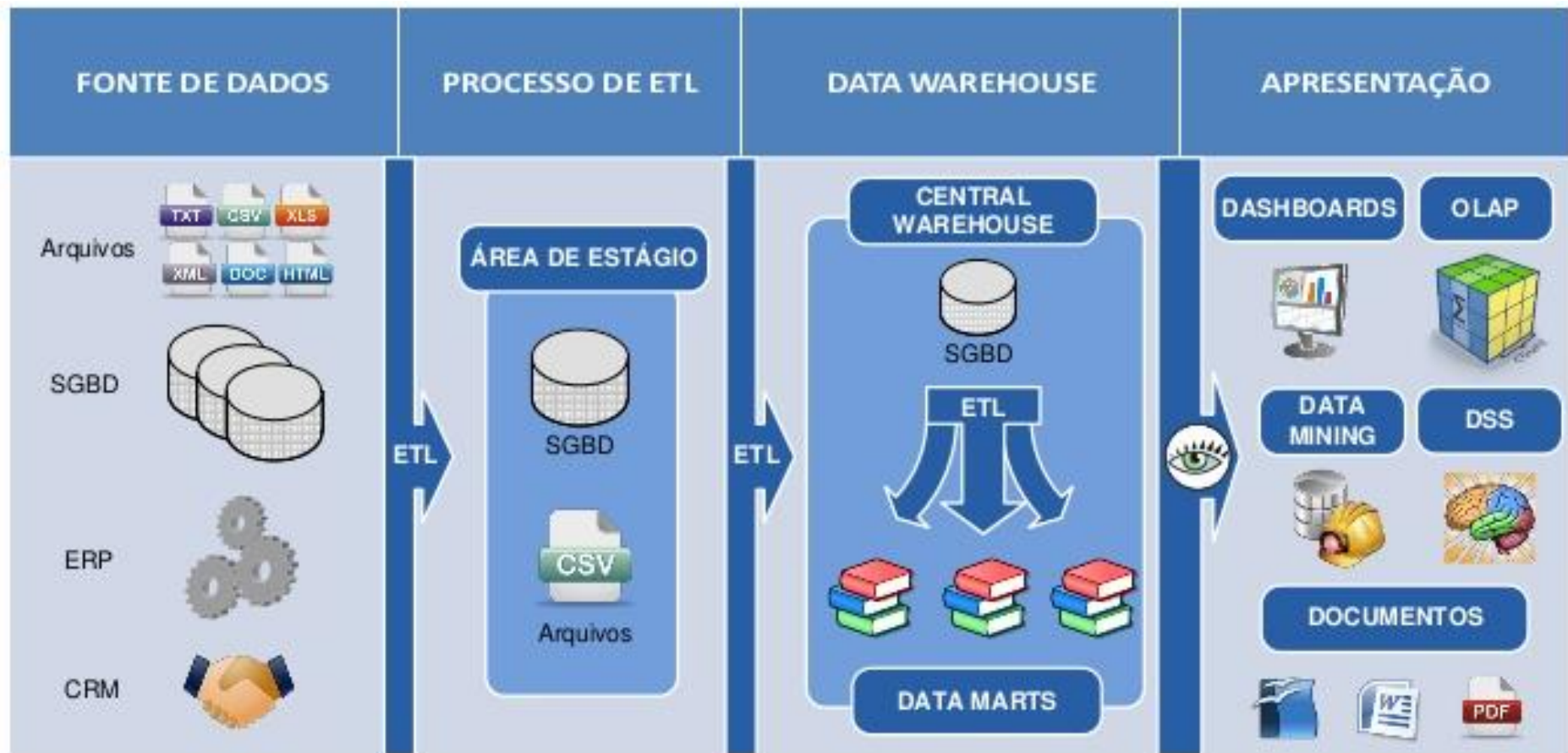


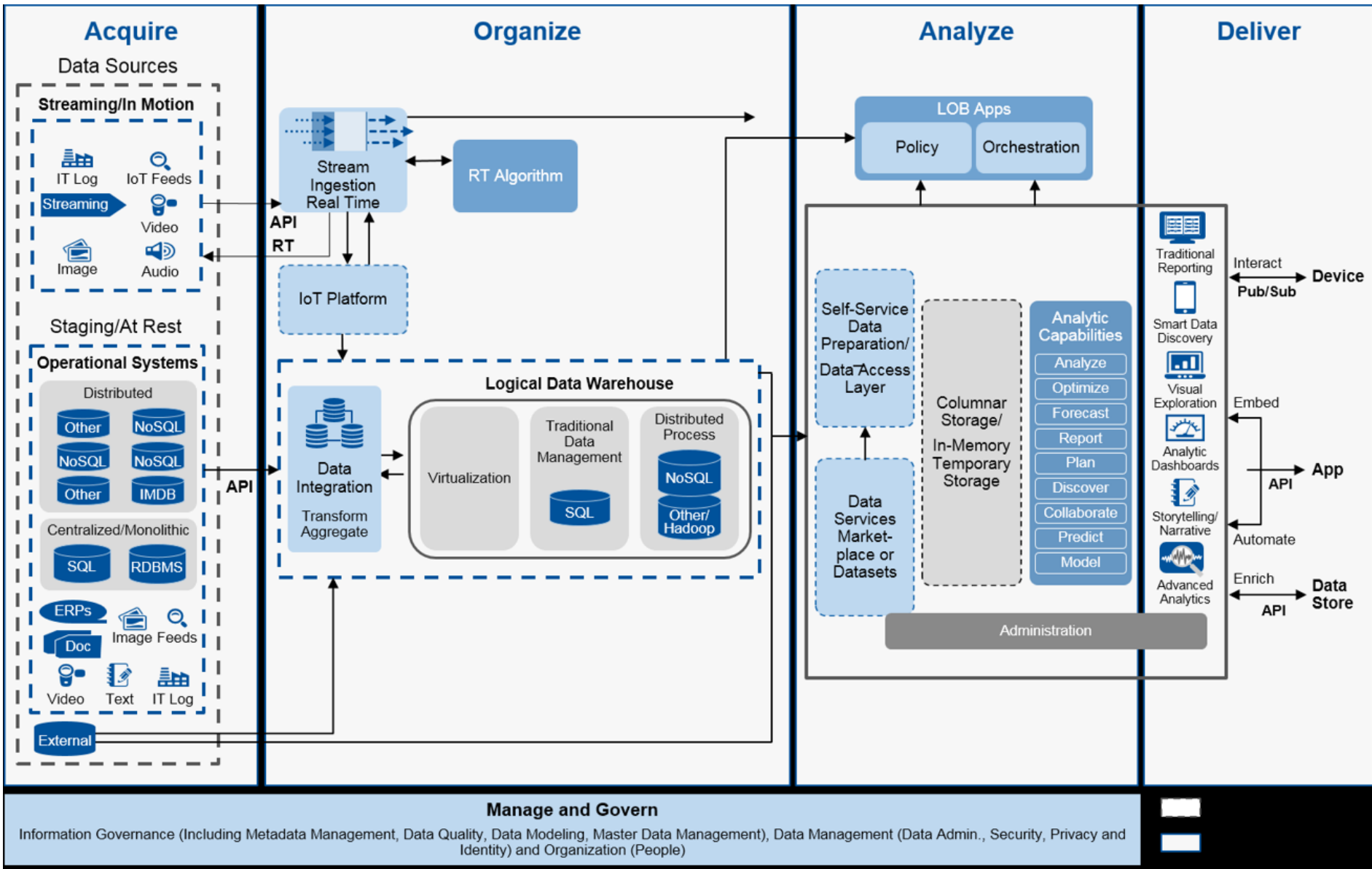
Como Criar e Utilizar a Inteligência



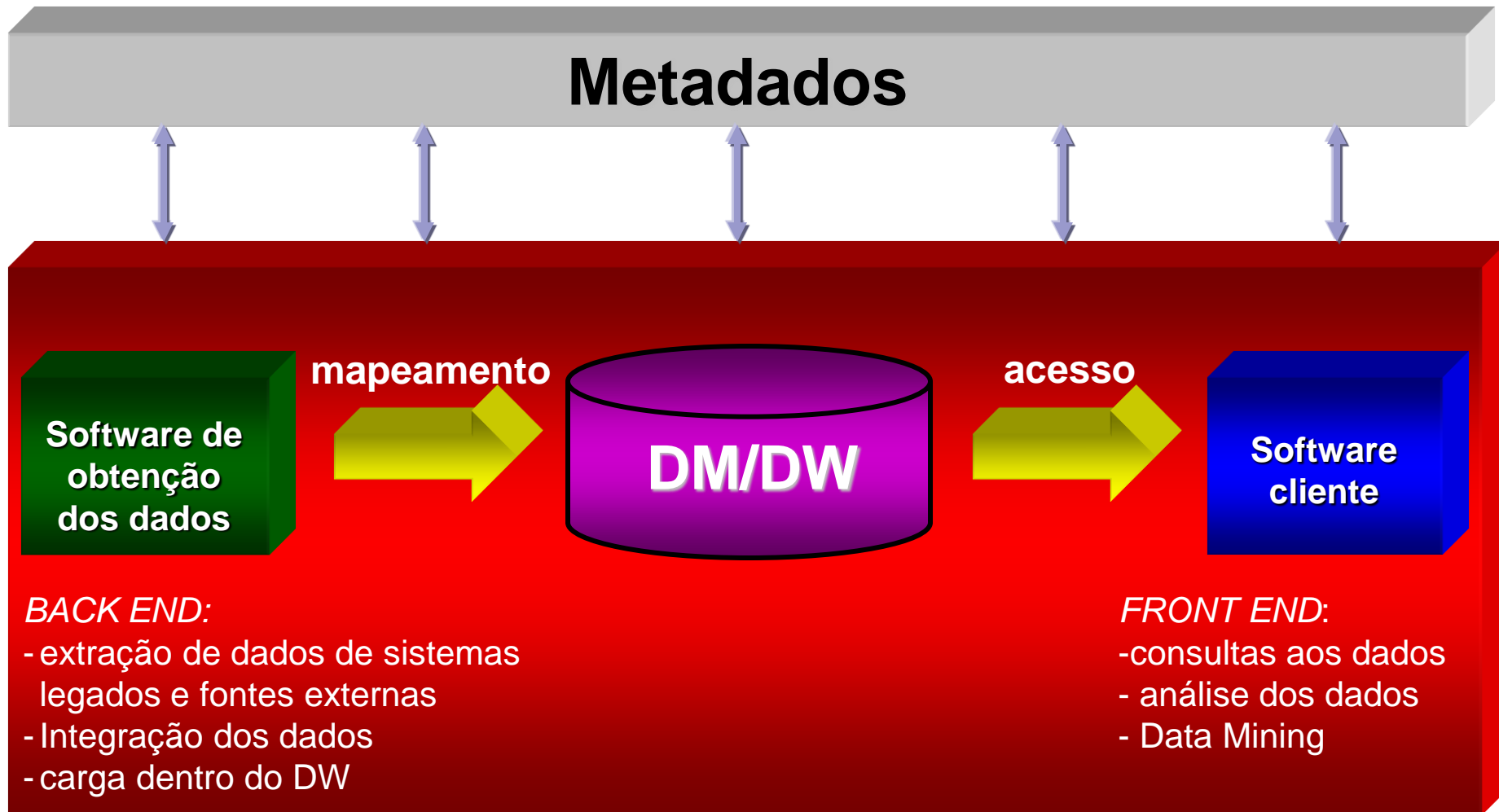
1. Análise
2. Produção
3. Disseminação
4. Uso
5. Exigência/Feedback
6. Planejamento e atribuição das tarefas
7. Coleta
8. Processamento e Exploração

Arquitetura Genérica da Plataforma BI





Principais Componentes do BI

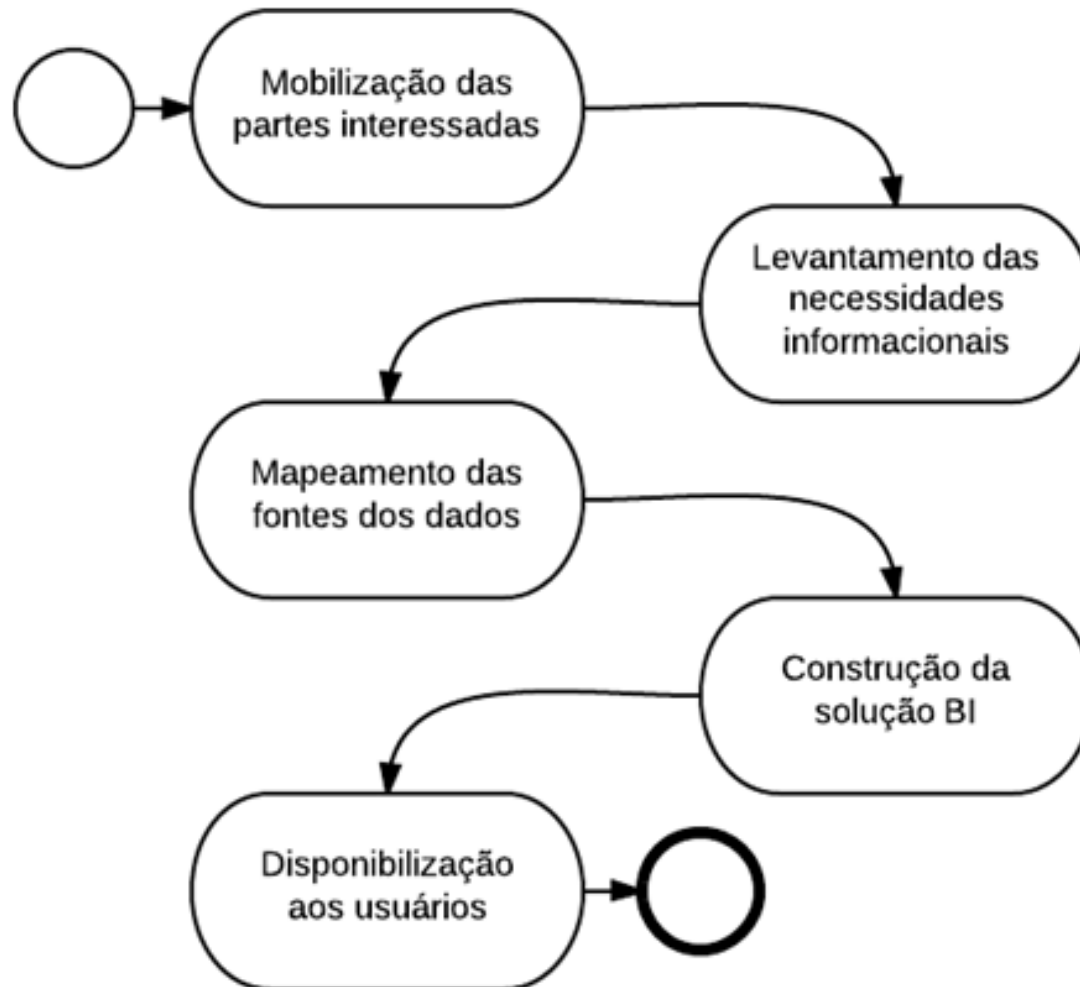




Processo de Desenvolvimento



Processo de Desenvolvimento

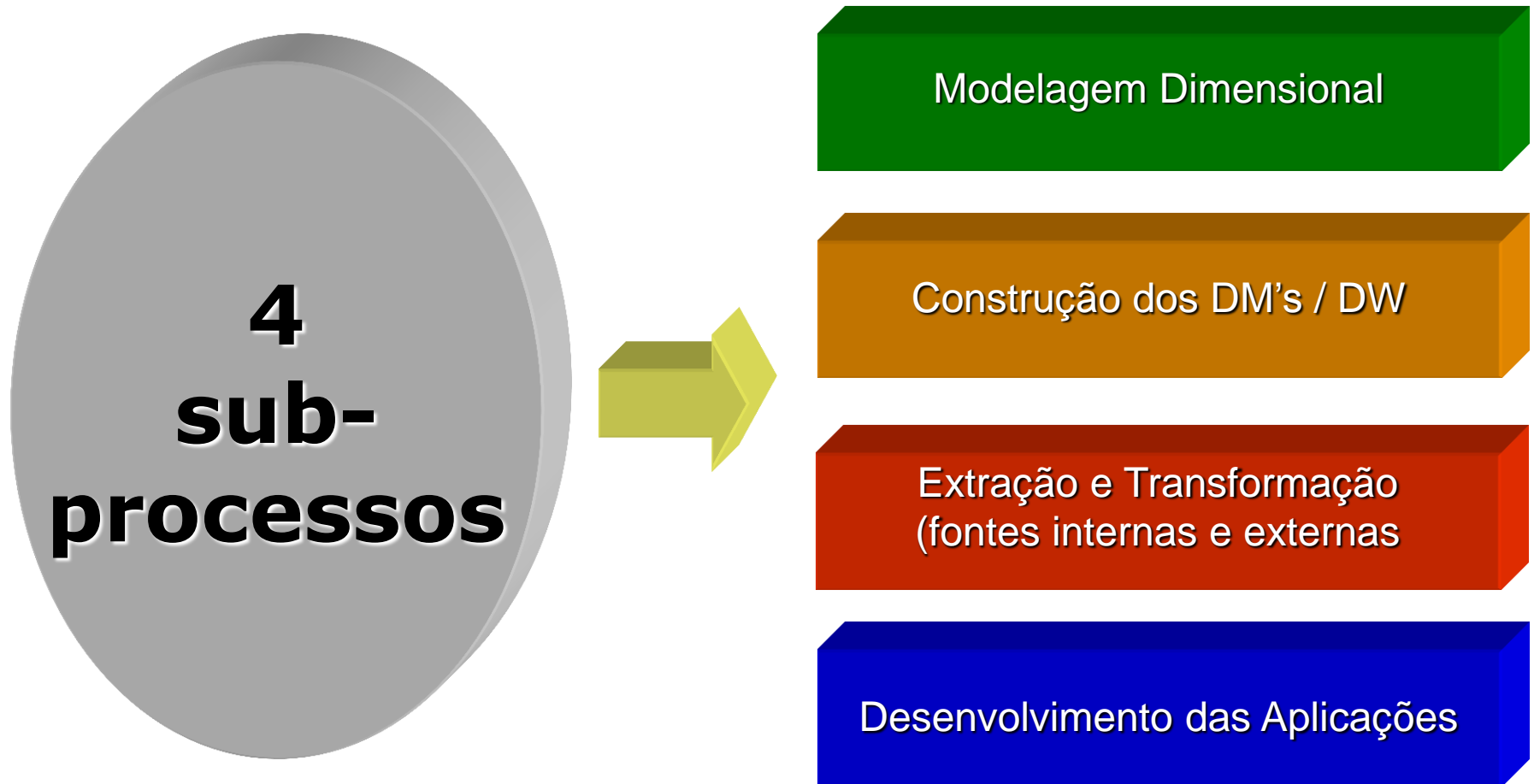


Processo de Desenvolvimento



KickOff	Planejamento	Indicadores	Modelo Lógico	Modelo Físico	Desenvolvimento
Avaliação Inicial	Análise situação atual	Análise indicadores existentes	Identificação das métricas	Mapeamento de fontes de dados	Desenvolvimento
Consenso sobre a necessidade	Coleta de materiais	Revisão dos indicadores	Identificação das dimensões	Avaliação de Plataforma de BI	Gestão de projeto
Entendimento de conceitos	Análise processo gestão	Avaliação de metas	Mapeamento de processos	Critérios de Data Quality	
Definir Sponsor	Definição key users e equipe	Detalhamento dos indicadores	Mapeamento de sistemas e funcionalidades	Estratégia de Atualização	
Identificar benefícios	Definição escopo	Workshop de revisão	Nível de detalhamento por métrica		
Aprovação Metodologia			Contextualização de métricas e atributos		
Entregáveis:	Entregáveis:	Entregáveis:	Entregáveis:	Entregáveis:	Entregáveis:
Workshop com líderes	Plano de desenvolvimento	Indicadores e conceitos	Modelo lógico	Modelo físico dos dados	Escopo desenvolvido
	Macro cronograma		Cronograma detalhado	Dicionário de Dados	Treinamento
			Escopo definido		

Processo de Desenvolvimento



Pontos de Atenção



- Integração de dados e metadados de várias fontes
- Qualidade dos dados: limpeza e refinamentos
- Sumarização e agregação de dados
- Sincronização das fontes com o *data warehouse* para assegurar atualidade
- Problemas de desempenho relacionados ao compartilhamento do mesmo ambiente computacional para abrigar os BDs corporativos operacionais e o *data warehouse*.

Fatores de Sucesso!



- Compromisso da alta administração
- Um time dedicado de analistas de BI
- Um sólido modelo de dados (DW/DM/LDW)
- Um plano de implementação bem elaborado
- Administração dinâmica do sistema
- Ferramenta de acesso aos dados de fácil uso e intuitiva pelos usuários finais

Data Warehouse

DW & BI





Data Warehouse

Conceitos

“É a separação física dos sistemas de dados operacionais de uma organização, de seus sistemas de suporte à decisão”. *Harry Singh*

“É um banco de dados, voltado para suporte à decisão, não volátil, variante no tempo e orientado a assuntos”. *Bill Inmon*

“Não consiste apenas de dados, mas também em um conjunto de ferramentas para consultar, analisar e apresentar informações”. *Ralph Kimball*



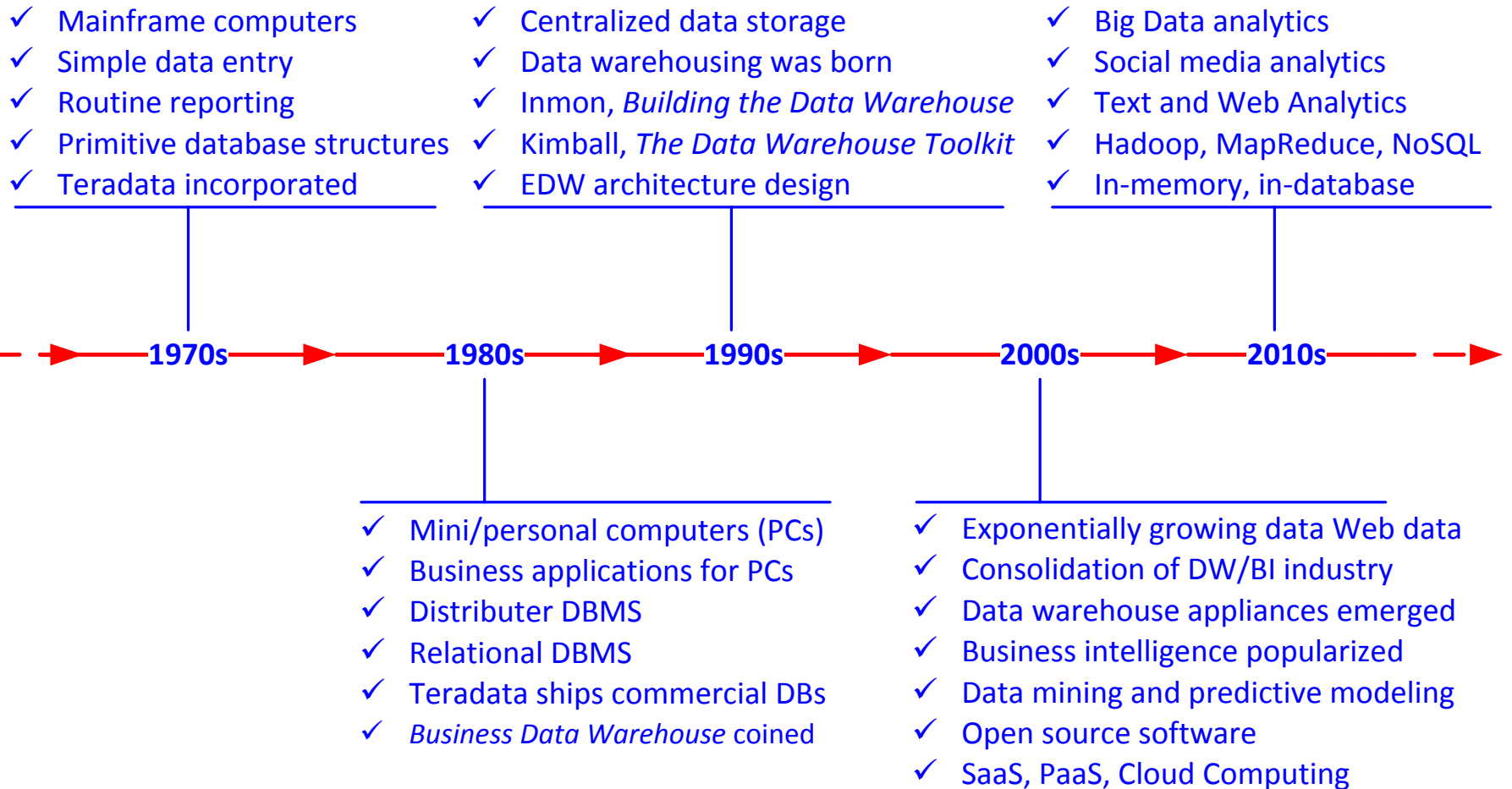
Data Warehouse

Histórico

- Criado pela IBM na década de 60 com o nome de *Information Warehouse*
- Tornou-se viável com o surgimento de novas tecnologias para armazenar e processar uma grande quantidade de dados
- O nome Data Warehouse foi dado por Willian H. Inmon, considerado o pai dessa tecnologia, em 1990
- Aproximadamente em 1994 ficou mais conhecido

Data Warehouse

Histórico





Data Warehouse

Para que server?

- Armazenar dados históricos usados no processo de tomada de decisão
- Criar uma visão única e centralizada dos dados que estavam dispersos em diversos bancos de dados
- Permitir que os usuários finais executem consultas, gerem relatórios e façam análises



Data Warehouse

Principais Características

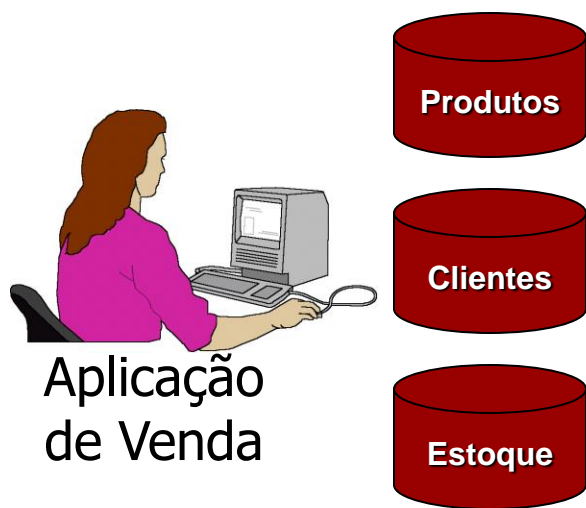
- De acordo com a definição de Inmon, um DW deve ser:
 - Orientado a assunto
 - Integrado
 - Não-volátil
 - Variável com o tempo

Data Warehouse

Principais Características

■ Orientação a assunto

- As informações são organizadas de modo a facilitar a análise dos dados, por exemplo: vendas, marketing, etc
- Os dados são organizados por assunto e não por aplicação, como em BDs operacionais

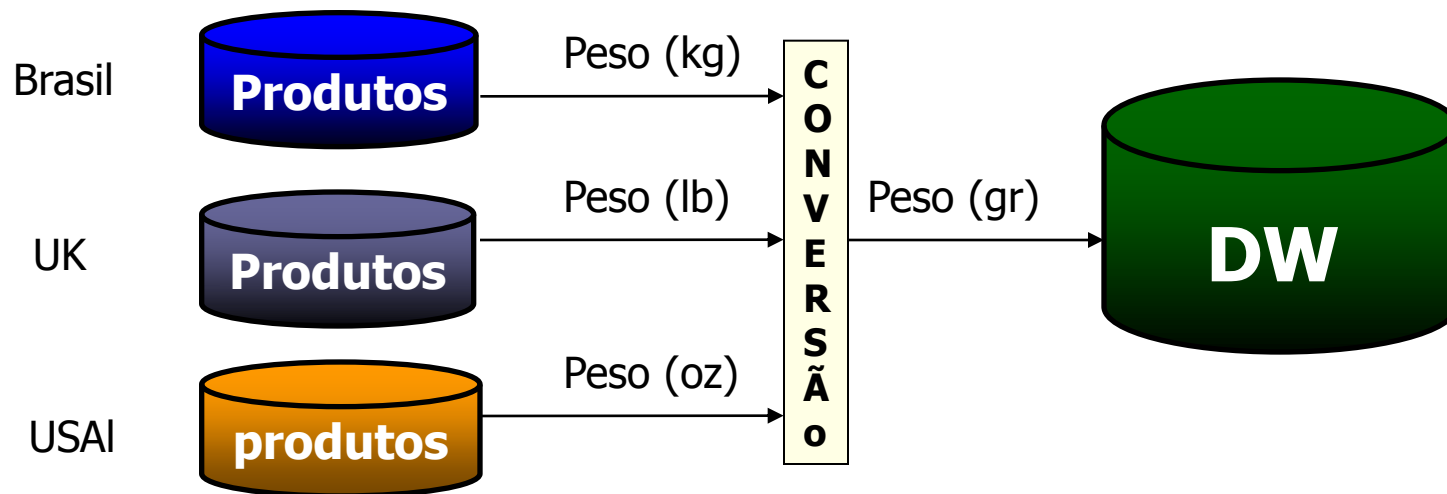


Data Warehouse

Principais Características

■ Integração

- Dados do DW provêm de diversas fontes
- Dados podem ser sumarizados ou eliminados
- Formato dos dados deve ser padronizado para uniformizar nomes, unidades de medidas, etc.

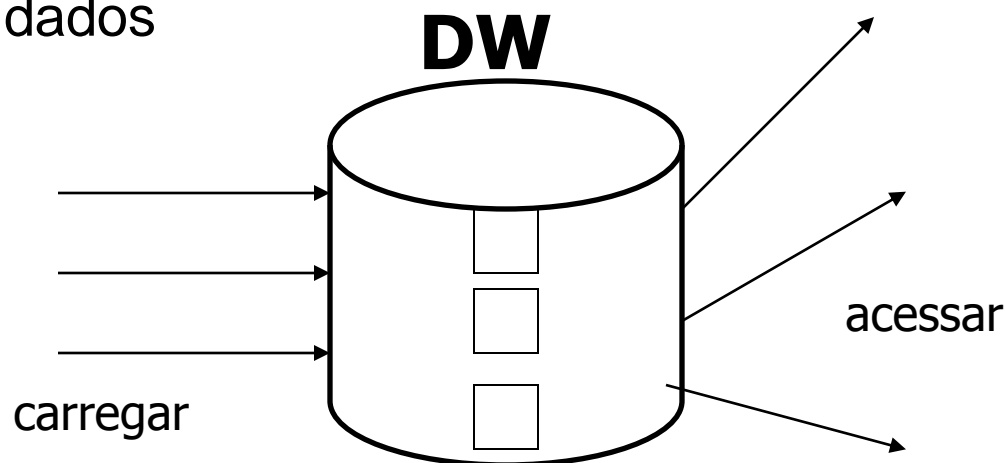


Data Warehouse

Principais Características

■ Não-volátil

- Uma vez que corretamente gravados no DW, os dados não podem ser alterados ou eliminados
- Com isso, garante-se que consultas subsequentes a um dado produzirão o mesmo resultado
- As operações possíveis são as de carga dos dados e de acesso aos dados



Data Warehouse

Principais Características

- Variável com o tempo

- Os dados no DW são relativos a um determinado instante de tempo



Produto	Preço
Caneta	0,50
Lápis preto	0,30
...	...



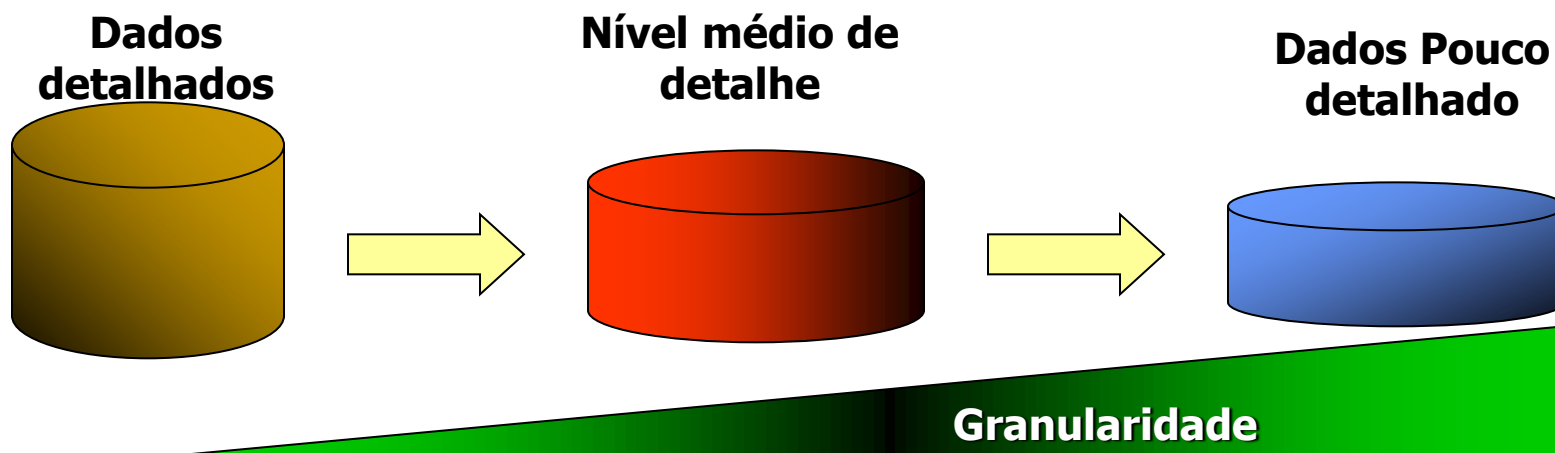
Produto	Jan/2017	Fev/2017	Mar/2017
Caneta	0,40	0,45	0,50
Lápis preto	0,25	0,28	0,30
...

Data Warehouse

Principais Características

■ Granularidade

- É o nível de detalhe, ou sumarização, ou resumo dos dados no DW
- O nível de granularidade
- É de extrema importância no projeto DW
 - Tem efeito direto no tamanho do BD e no tipo de análise que o BD pode suportar





Data Warehouse

Granularidade

- Definir a granularidade adequada é vital para que o DW atenda aos seus objetivos
 - **Menor Granularidade:**
 - Mais detalhes → Mais dados → Análises mais longas → Informação mais detalhada
 - **Maior Granularidade:**
 - Menos detalhes → Menos dados → Análises mais curtas → Informação menos detalhada
- Normalmente, para evitar que se perca informação, são criados vários níveis de granularidade



Data Warehouse

Granularidade

- Procurar um equilíbrio entre o gerenciamento do volume dos dados e a armazenagem dos dados
 - Cuidado com um nível muito alto que impossibilite a utilização detalhada
- Passos:
 - Usar bom senso → criar uma pequena parte de DW, deixar o usuário acessar os dados, colher feedback
 - Estimar o número de linhas de dados e do dispositivo de armazenamento.
 - Planejar → quanto espaço de disco é necessário e se são necessários mais de um nível de granularidade

Data Warehouse

Granularidade

- Exemplo: rastreamento de compras com cartão de crédito
 - 3 anos -> 36 meses
 - número de contas -> 50 milhões
 - número médio de compras por mês/conta -> 30
 - número de registros -> $36 \times 50.000.000 \times 30 = 54$ bilhões
 - tamanho do registro -> 8 campos de 4 bytes
 - tamanho básico da tabela -> $54 \text{ bilhões} \times 8 \times 4 = 1730 \text{ Gb} = 1.73 \text{ TB}$

[Inmon]



Data Warehouse

Granularidade

■ Processo de sumarização

- Aplica um novo esquema de modo a condensar os dados
 - ✓ Ex.: armazenar totais, médias, etc.

■ Processo de envelhecimento

- Transfere os dados antigos do HD para fita, CD, etc.
- Mantém o nível de detalhe para que nenhuma informação seja perdida

Data Warehouse

Granularidade – Exemplo Cenário Bancário

Operacional – 60 dias

- número conta
- data ocorrência
 - valor
 - caixa
 - local
 - identificação (débito/crédito)
 - para quem
 - saldo da conta

DW – até 10 anos

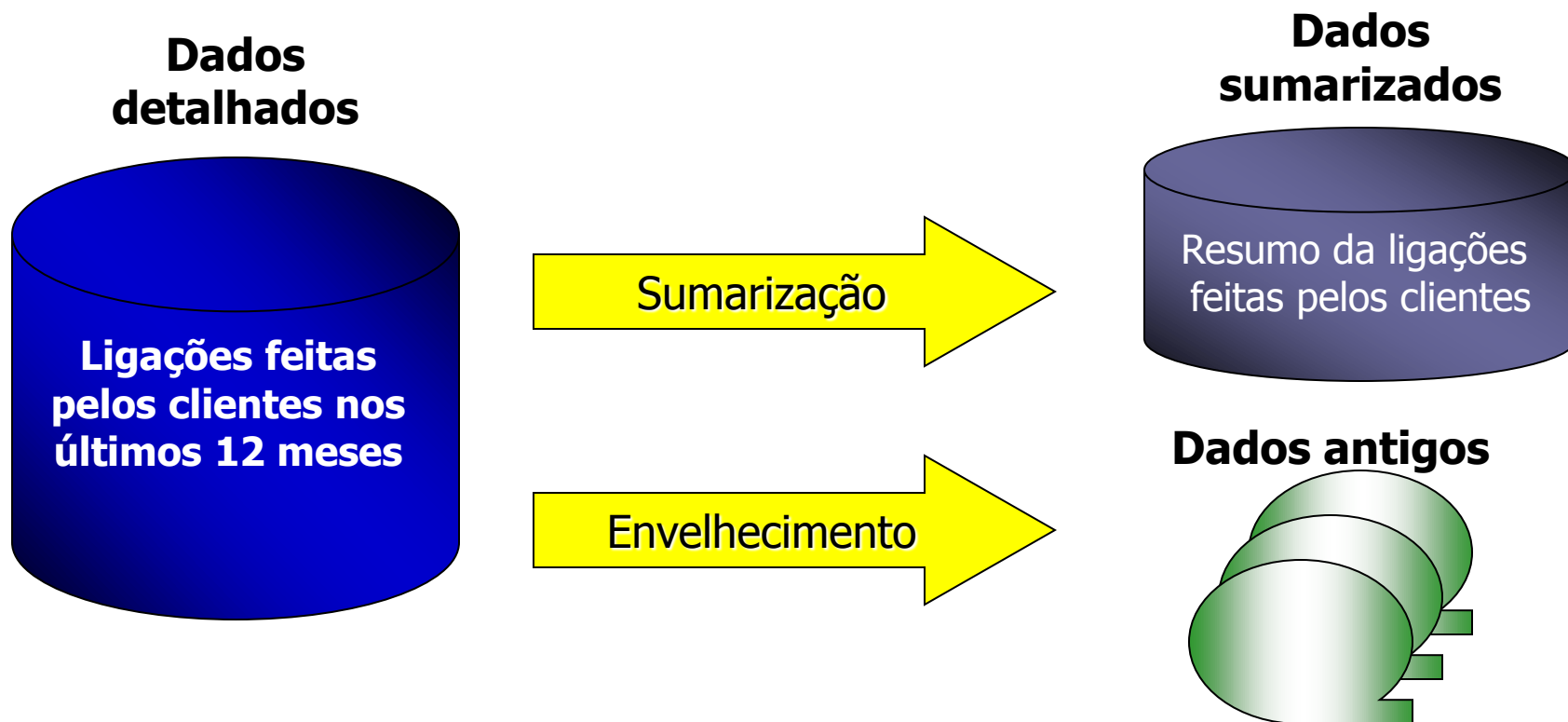
- número conta
- mês
 - número de transações
 - saques
 - depósitos
 - saldo inicial
 - saldo final
 - maior saldo da conta
 - menor saldo da conta
 - saldo médio da conta

DW – até 10 anos

- número conta
- data ocorrência
 - valor
 - identificação (débito/crédito)
 - saldo da conta

Data Warehouse

Granularidade – Exemplo Cenário Empresa Telefônica



Data Warehouse

Granularidade – Exemplo Cenário Empresa Telefônica

Dados Detalhados

Ligações
Origem
Destino
Início
Fim
Tarifa
Status

Num. Registros:
Ligações nos últimos
12 meses

Dados Sumarizados

Ligações
Cliente
Mês
Pulsos
Valor conta
Longa dist.
Vencimento

Num. Registros:
Contas emitidas
pela empresa

Dados Antigos

Ligações
Origem
Destino
Início
Fim
Tarifa
Status

Num. Registros:
Ligações efetuadas
pela empresa

Data Warehouse

Granularidade – Exemplo Cenário Empresa Telefônica

- Quanto menor a granularidade, mais detalhada é a informação disponível
 - No exemplo anterior, poderíamos determinar se o cliente A ligou para B na semana passada
 - Também poderíamos verificar se A faz muitas chamadas de longa distância

Dados Detalhados

Ligações

Origem

Destino

Início

Fim

Tarifa

Status

Data Warehouse

Granularidade – Exemplo Cenário Empresa Telefônica

- Durante o processo de sumarização, algumas informações podem ser perdidas
 - Não seria possível saber se A ligou para B
 - É possível verificar o padrão de consumo de A

Dados Sumarizados

Ligações

Cliente

Mês

Pulsos

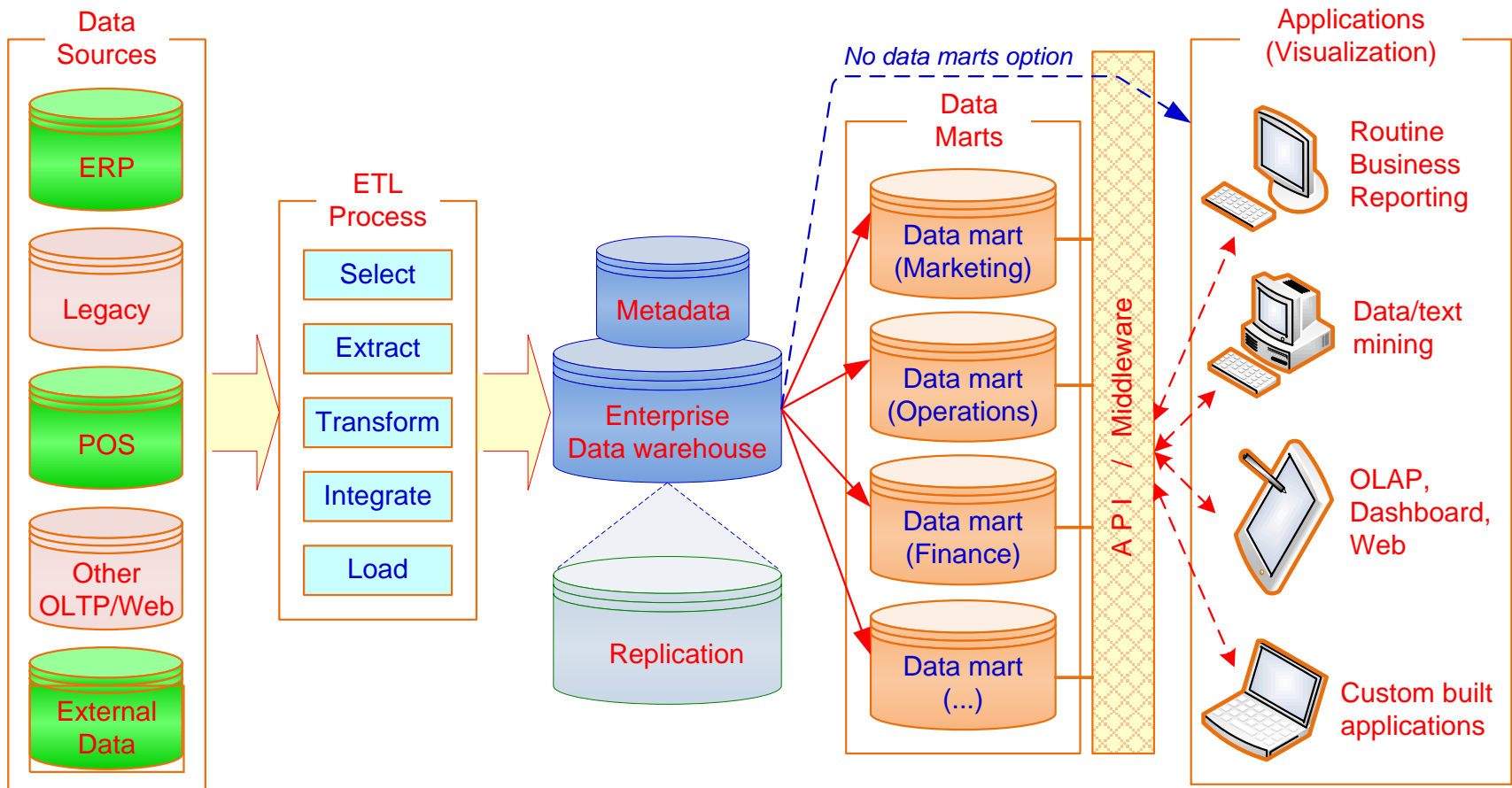
Valor conta

Longa dist.

Vencimento

Data Warehouse

Arquitetura Genérica



Data Warehouse

Arquiteturas

■ 3 Camadas

- Software de aquisição de dados (back-end)
- O data warehouse contém dados e o software
- Software cliente (front-end) permite os usuários acessar e analisar dados do *data warehouse*

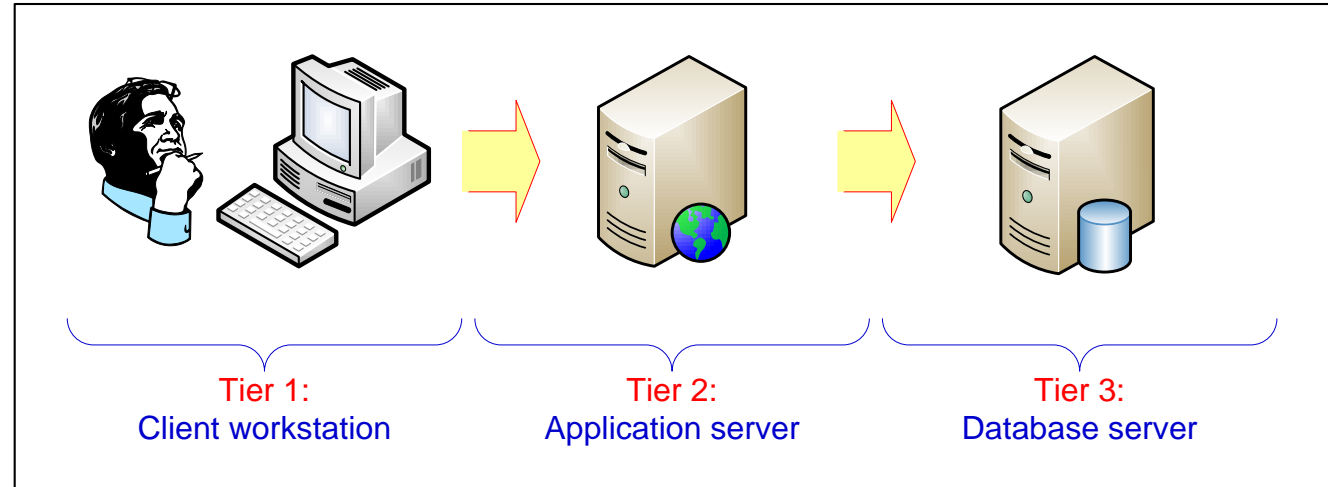
■ 2 Camadas

- As duas primeiras camadas na arquitetura de três camadas são combinadas em uma
- ... às vezes há apenas uma camada?

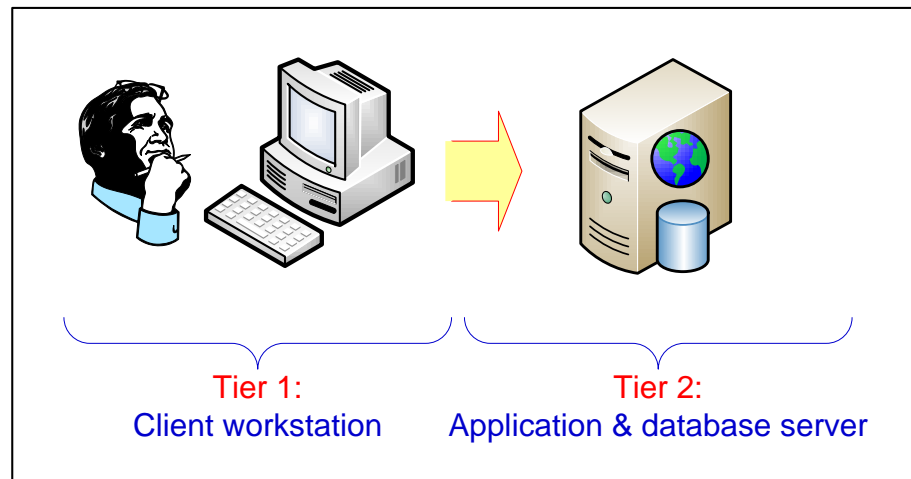
Data Warehouse

Arquiteturas

3-tier architecture



2-tier architecture



1-tier
Architecture
?



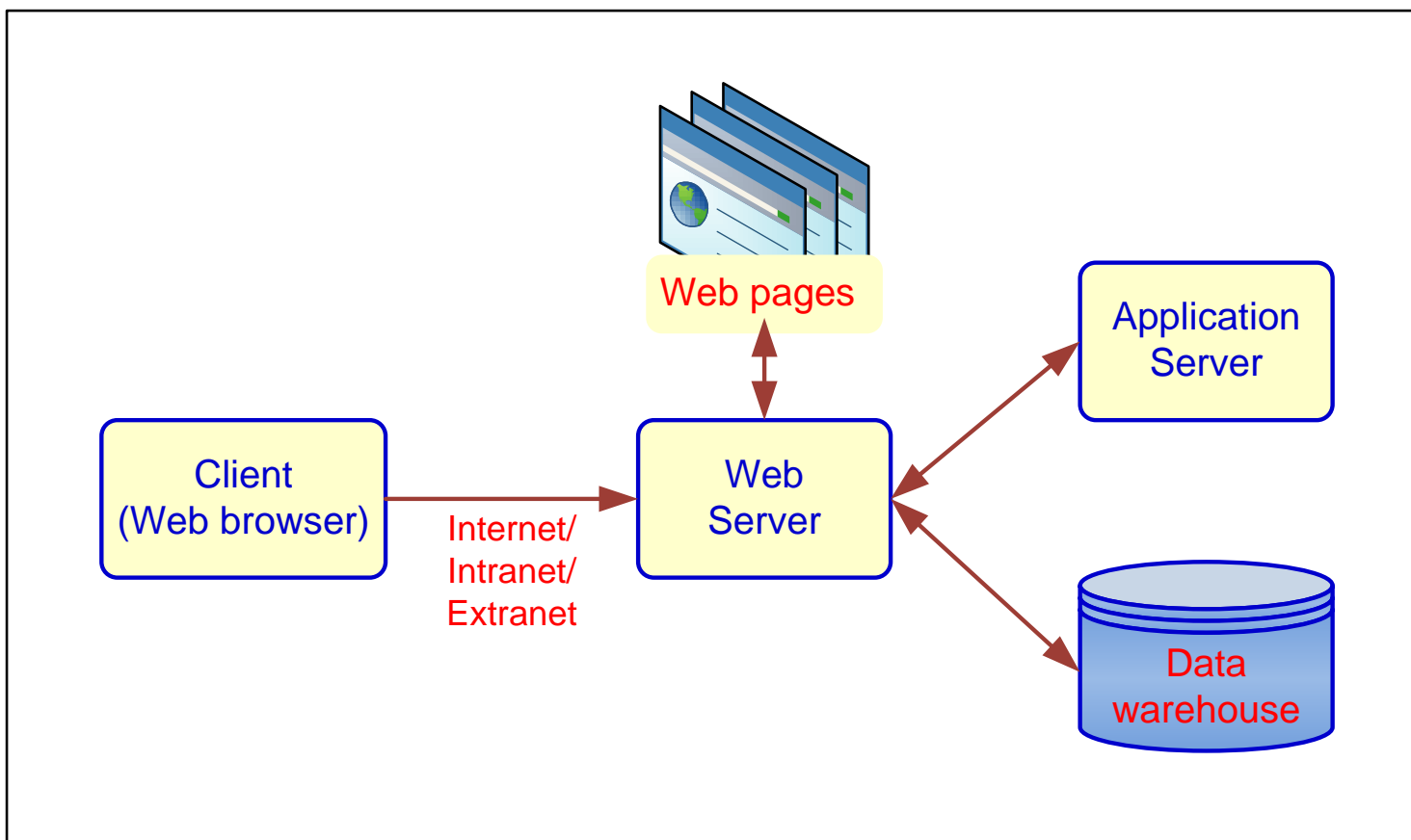
Data Warehouse

Arquiteturas

- Questões a serem consideradas ao decidir qual arquitetura usar:
 - Qual sistema de gerenciamento de banco de dados (DBMS) deve ser usado?
 - O processamento será paralelo e/ou particionamento será utilizado?
 - As ferramentas de migração de dados serão usadas para carregar o data warehouse?
 - Que ferramentas serão usadas para apoiar a recuperação e análise de dados?

Data Warehouse

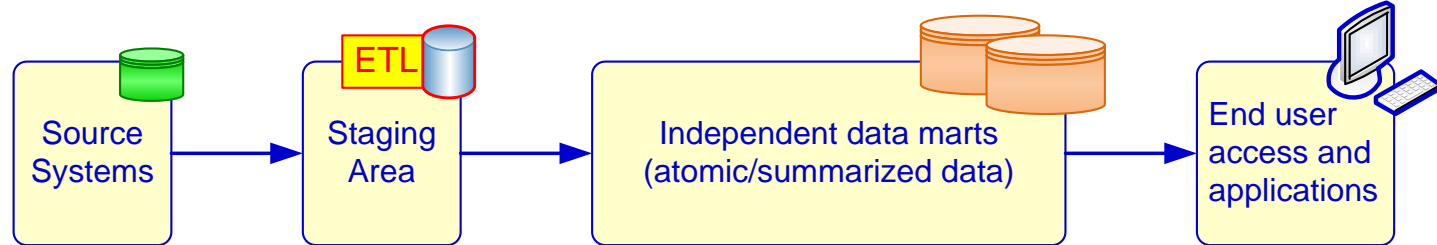
Arquitetura baseada na Web



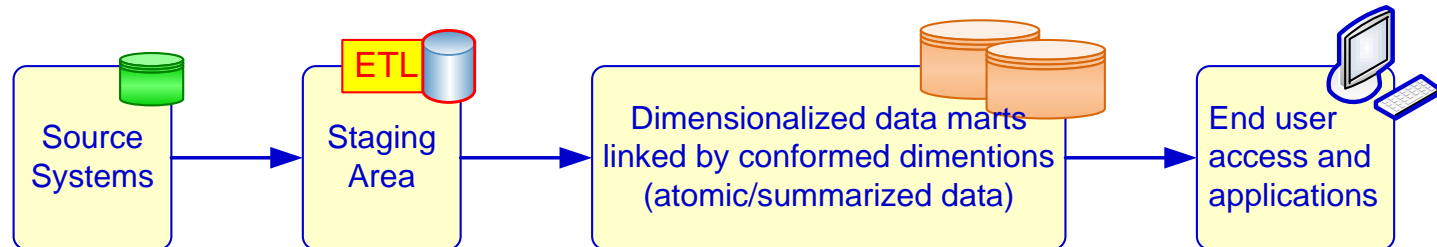
Data Warehouse

Arquiteturas Alternativas

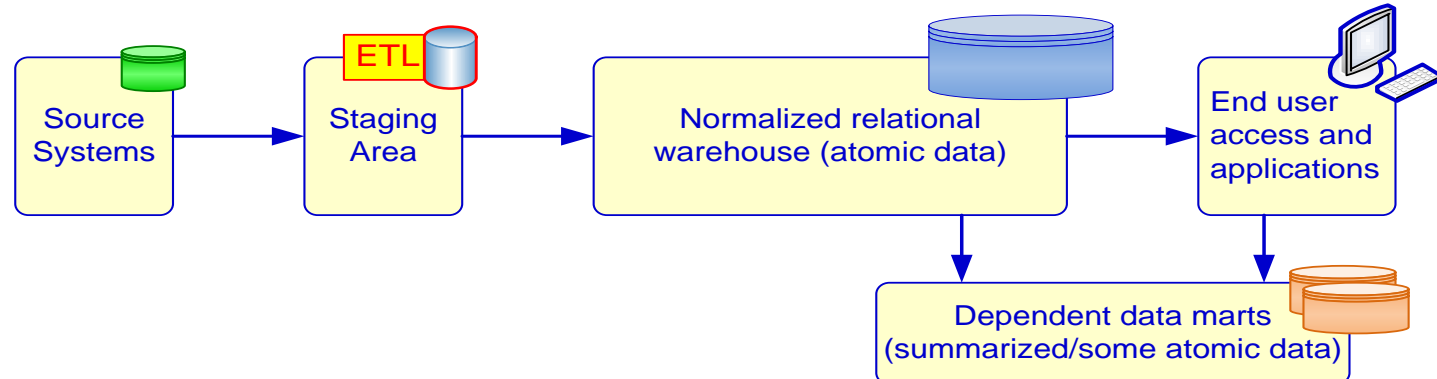
(a) Independent Data Marts Architecture



(b) Data Mart Bus Architecture with Linked Dimensional Datamarts



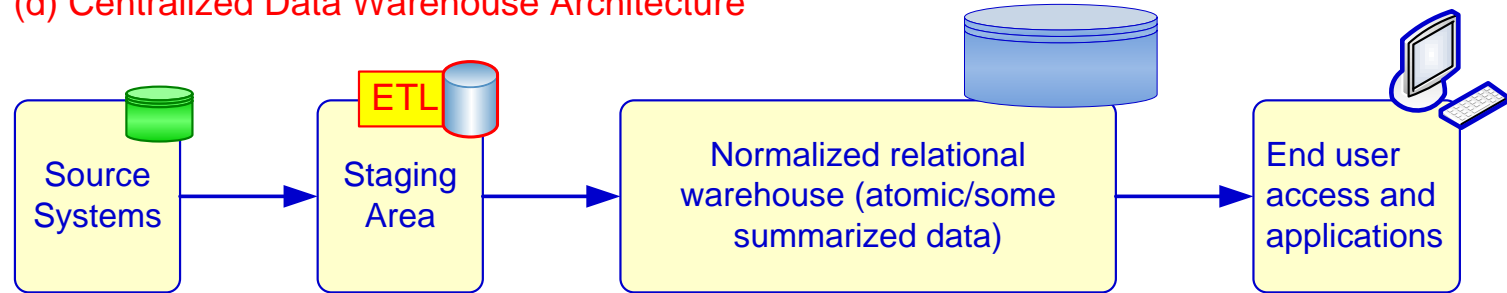
(c) Hub and Spoke Architecture (Corporate Information Factory)



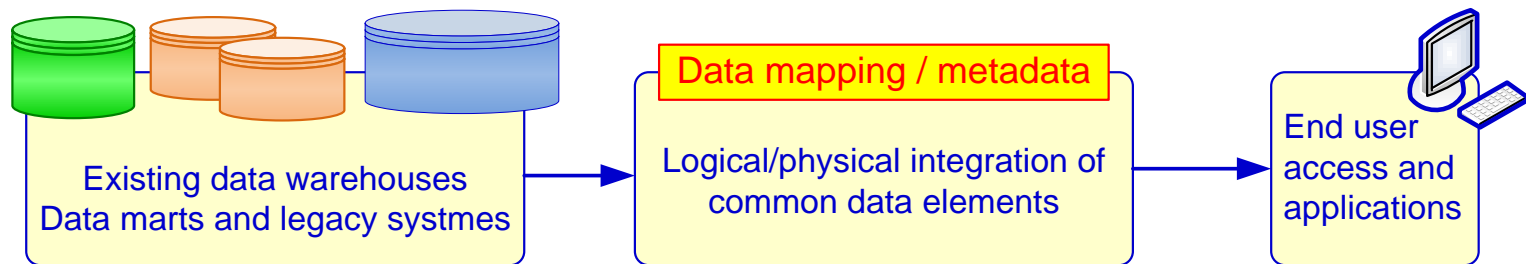
Data Warehouse

Arquiteturas Alternativas


(d) Centralized Data Warehouse Architecture



(e) Federated Architecture



- Cada arquitetura tem vantagens e desvantagens!
- Qual arquitetura é a melhor?



Dez fatores que potencialmente afetam a decisão de seleção de arquitetura DW

1. Interdependência de informação entre unidades organizacionais
2. Necessidades de informação da alta administração
3. Urgência da necessidade de um data warehouse
4. Natureza das tarefas do utilizador final
5. Restrições sobre os recursos
6. Visão estratégica do data warehouse antes da implementação
7. Compatibilidade com sistemas existentes
8. Percepção da capacidade da equipe de TI
9. Problemas técnicos
10. Fatores sociais / políticos

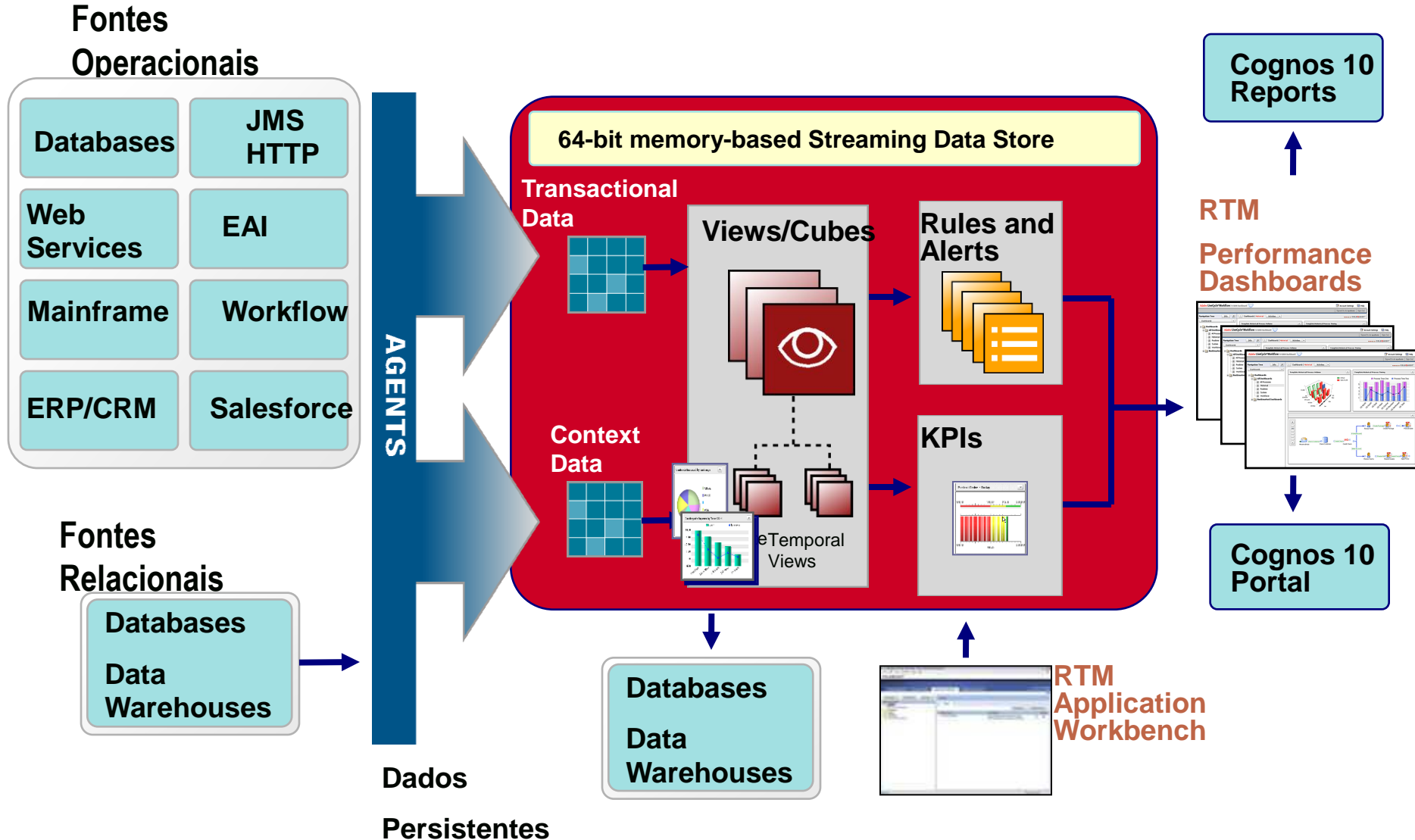


Data Warehouse

Outros Componentes

- *Operational data stores (ODS)*
 - Um tipo de banco de dados usado frequentemente como uma área intermediária para um data warehouse.
- *Oper marts*
 - Um data mart operacional.
- *Enterprise data warehouse (EDW)*
 - Um data warehouse corporativo.
- Metadata – “*data about data*”
 - Em DW, os metadados descrevem o conteúdo de um data warehouse e sua aquisição e uso.

Arquitetura *Real Time* – IBM Cognos RTM (Exemplo)



BD Operacional x *Data Warehouse*

	BD operacional	Data Warehouse
Usuários	Funcionários	Alta administração
Utilização	Tarefas cotidianas	Decisões estratégicas
Padrão de uso	Previsível	Difícil de prever
Princípio de funcionamento	Baseado em transações	Baseado em análise de dados
Valores dos dados	Atuais e voláteis 60 a 90 dias	Históricos e imutáveis 5 a 10 anos
Detalhamento	Alto	Baixo e sumarizado
Organização dos dados	Orientado a aplicações OLTP	Orientado a assunto OLAP
Volume dados	Gigabytes – Terabytes	Terabytes - Pentabytes

A Evolução do Data Warehouse

■ *Fontes de dados...*

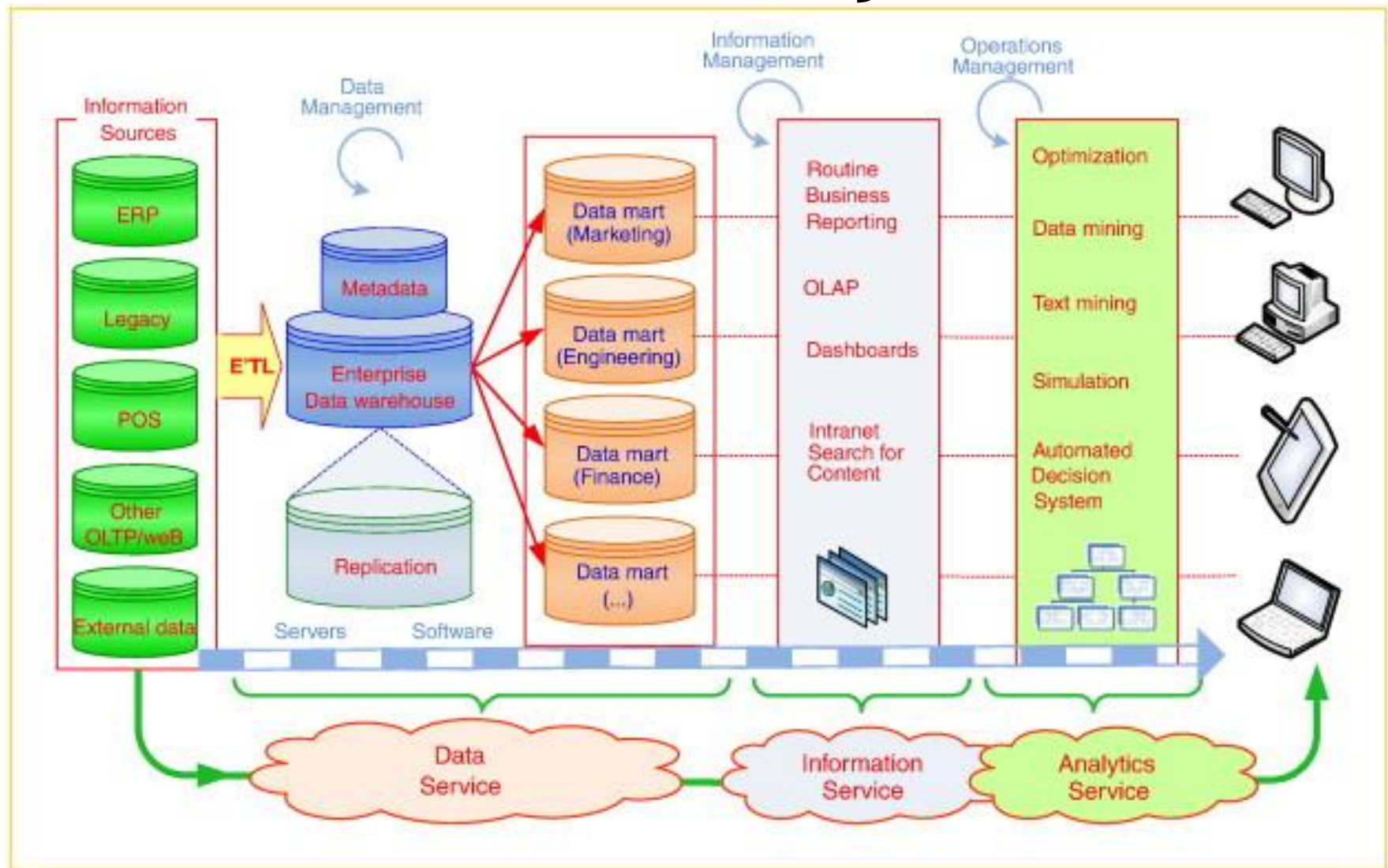
- Web, social media, and Big Data
- Open source software
- SaaS (software as a service)
- Cloud computing
- Data lakes

■ *Infraestruturas...*

- Columnar
- Real-time DW
- Data warehouse appliances
- Data management practices/technologies
- In-database & In-memory processing New DBMS
- New DBMS, Advanced analytics, ...



BI Orientado a Serviço



BD Operacional x *Data Warehouse*

	BD operacional	Data Warehouse
Usuários	Funcionários	Alta administração
Utilização	Tarefas cotidianas	Decisões estratégicas
Padrão de uso	Previsível	Difícil de prever
Princípio de funcionamento	Baseado em transações	Baseado em análise de dados
Valores dos dados	Atuais e voláteis 60 a 90 dias	Históricos e imutáveis 5 a 10 anos
Detalhamento	Alto	Baixo e sumarizado
Organização dos dados	Orientado a aplicações OLTP	Orientado a assunto OLAP
Volume dados	Gigabytes – Terabytes	Terabytes - Petabytes

Especialização em Ciência de Dados com Big Data, BI e *Data Analytics*



Prof. Fernando Siqueira
fssiqueira@gmail.com