



Business Intelligence

DATA WAREHOUSE

Business Intelligence - BI

Conteúdo:

- 1) Business Intelligence
- 2) Data Warehouse / Data Mart / Data Lake
- 3) Modelagem Dimensional
- 4) ELTs
- 5) Dimensões alteradas lentamente
- 6) Ferramentas OLAP

Business Intelligence - BI

Conteúdo:

1) Data Warehouse – DW:

- Introdução
- Características
- Qualidade dos dados
- Elementos
- Data Mart



1 - Introdução

CONHECENDO O DATA WAREHOUSE

2 - Data warehouse - DW

Introdução

- Criado pela IBM na década de 60.
- O nome DATA WAREHOUSE foi dado por WILLIAM H. INMON, considerado o pai desta tecnologia;
- Todos os dados da Organização deveriam estar armazenados em um único Banco de Dados que seria utilizado por todos os departamentos;

2 - Data warehouse - DW

Conceito

“É uma coleção de dados orientados por assuntos, integrados, variáveis no tempo e não voláteis, para dar suporte ao processo gerencial de tomada de decisão ”

(Inmon)

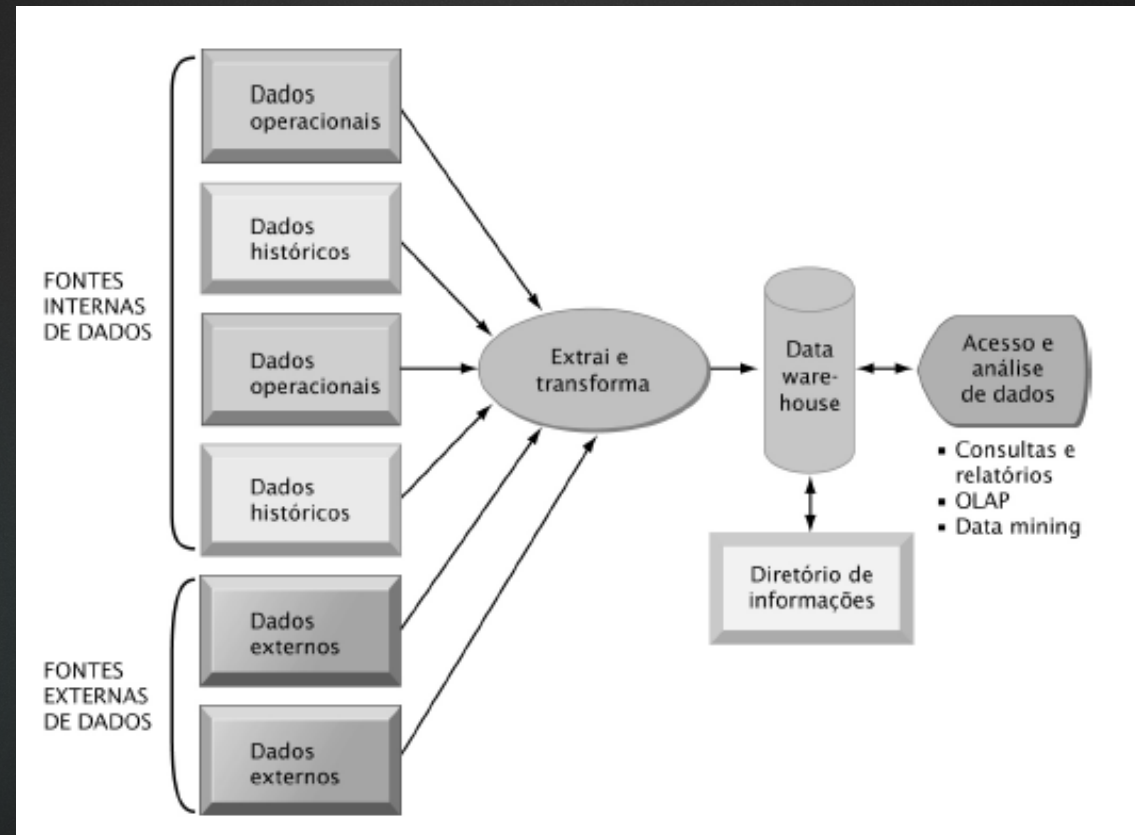
2 - Data warehouse - DW

Objetivo

- Fornecer dados para a geração de relatórios, análise de grandes volumes de dados e obtenção de informações estratégicas para apoiar a tomada de decisão;
- Por sua capacidade de sumarizar e analisar grandes volumes de dados, o DATA WAREHOUSE é o núcleo dos sistemas de informações gerenciais, e de apoio à decisão das principais soluções de Business Intelligence do mercado.

2 - Data warehouse - DW

Exemplo



2 - Data warehouse - DW

Data Warehouse x BD Operacional

Características	BD Operacional	Data warehouse
Objetivo	Operações diárias	Analisar o negócio
Uso	Operacional	Informativo
Tipo de processamento	OLTP	OLAP
Unidade de trabalho	Inclusão, alteração, exclusão	Carga e consulta
Tipo de usuários	Operadores	Gerencial
Condições dos dados	Dados operacionais	Dados Analíticos

2 - Data warehouse - DW

Onde usar

Características	BD Operacional
Vendas	Análise de vendas, Previsões, Lucratividade de Cliente/Contrato, análise de Canais de Distribuição
Manafatura	Análise de Estoque, Análise de Fornecedores, Planejamento de demanda, Análise de custos de matéria prima
Marketing	Determinação de Preço, Lucratividade de Produto, Análise de Oportunidades de Mercado
RH	Produtividade, Rendimento, Projeção de Salários, Análise de Headcount.

Características

DO DW

2 - Data warehouse - DW

Características

1. Orientação por assunto;
2. Integração;
3. Variação no tempo;
4. Não volatilidade;
5. Localização;
6. Credibilidade dos dados;
7. Granularidade;
8. Metadados.

2 - Data warehouse - DW

1-orientação por assunto

- Armazena dados importantes sobre temas específicos da empresa e conforme o interesse das pessoas que irão utilizá-lo;
- Exemplo:
 - Empresa trabalha com vendas de produtos no varejo e o seu maior interesse ser o perfil de seus compradores, então o DW será voltado para as pessoas que compram seus produtos e não para os produtos que ela vende.

2 - Data warehouse - DW

1-orientação por assunto

- Dados são organizados por assunto e não por aplicação, como em BDs operacionais:



Aplicação
de Venda



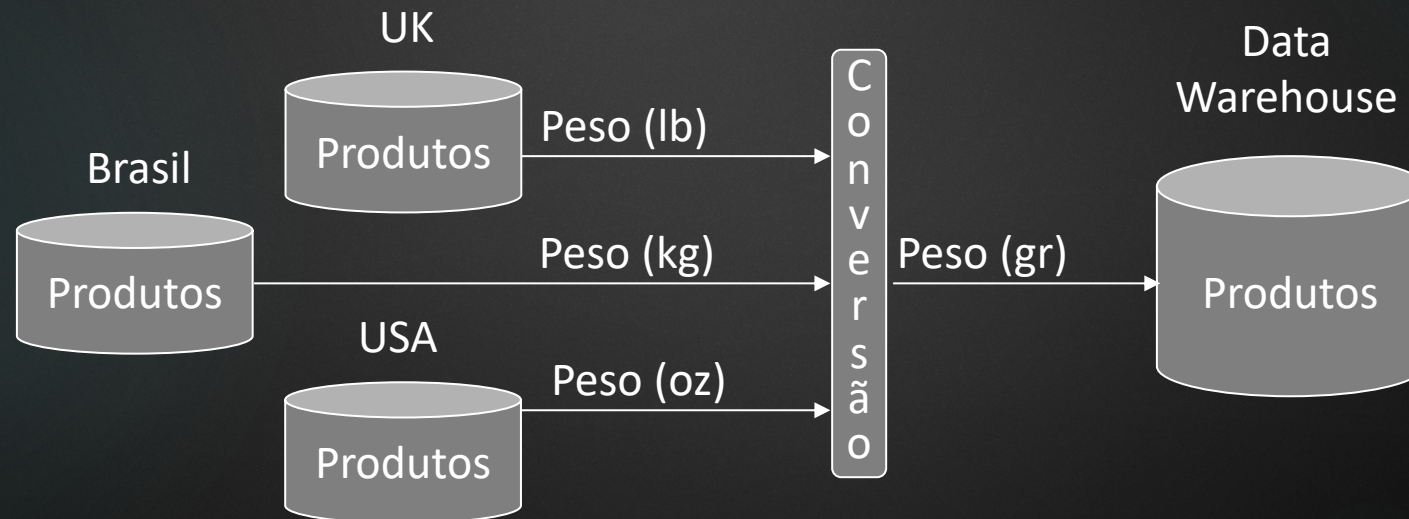
Análise
de Vendas



2 - Data warehouse - DW

2-Integração

- Dados de um DW provém de diversas fontes;
- Dados podem ser sumarizados ou eliminados;
- Formato dos dados deve ser padronizado para uniformizar nomes, unidades de medida, etc.



2 - Data warehouse - DW

2-Integração

Operacional	Data Warehouse
Gênero: Aplicação A: m, f Aplicação B: masculino, feminino	Gênero: M, F
Distância: Aplicação A: metros Aplicação B: pés	Distância: Centímetros
Chave PK ID: Aplicação A: char(10) Aplicação B: inteiro	Chave PK ID: Decimal

2 - Data warehouse - DW

2-Integração

Operacional - 3 Tabelas	Data Warehouse - 1 tabela
Plano de Saúde <ul style="list-style-type: none">- Maria Silva- Feminino- 01/12/1980	Cliente: <ul style="list-style-type: none">- Maria Silva- Feminino- Nascida em 01/12/1980- Duas internações em 2020- Equipe médica- Duração média das internações- Exames requeridos- Resultados dos exames
Clinica <ul style="list-style-type: none">- Maria Silva- Duas internações em 2020- Equipe médica- Duração média das Internações	
Laboratório de Exames <ul style="list-style-type: none">- Maria Silva- Exames requeridos- Resultados	

2 - Data warehouse - DW

3-Variação no tempo

- Dados não são mais alterados depois de incluídos no DW;
- Operações no DW:
 - Em um BD operacional é possível incluir, alterar e eliminar dados;
- Já no DW é possível apenas incluir dados;
- Garante que consultas subsequentes a um dado produzirão o mesmo resultado;

2 - Data warehouse - DW

4-Não volatilidade



2 - Data warehouse - DW

5-Granularidade

- O grão é o menor nível da informação e é definido de acordo com as necessidades levantadas no projeto;
- Determina o nível de detalhamento da informação;
- Quanto maior for o nível de detalhes, menor será o nível de granularidade
- O nível de granularidade afeta diretamente o volume de dados armazenados no DW
- Pode ser: Alta ou Baixa.

2 - Data warehouse - DW

5-Granularidade

- Baixa:
 - É possível responder a praticamente qualquer consulta;
 - Grande quantidade de recursos computacionais é necessária para responder perguntas específicas.
- Alta:
 - Ocorre uma significativa redução da possibilidade de utilização dos dados para atender consultas detalhadas;
 - Reduz muito o espaço em disco e o número de índices necessários.

2 - Data warehouse - DW

5-Granularidade

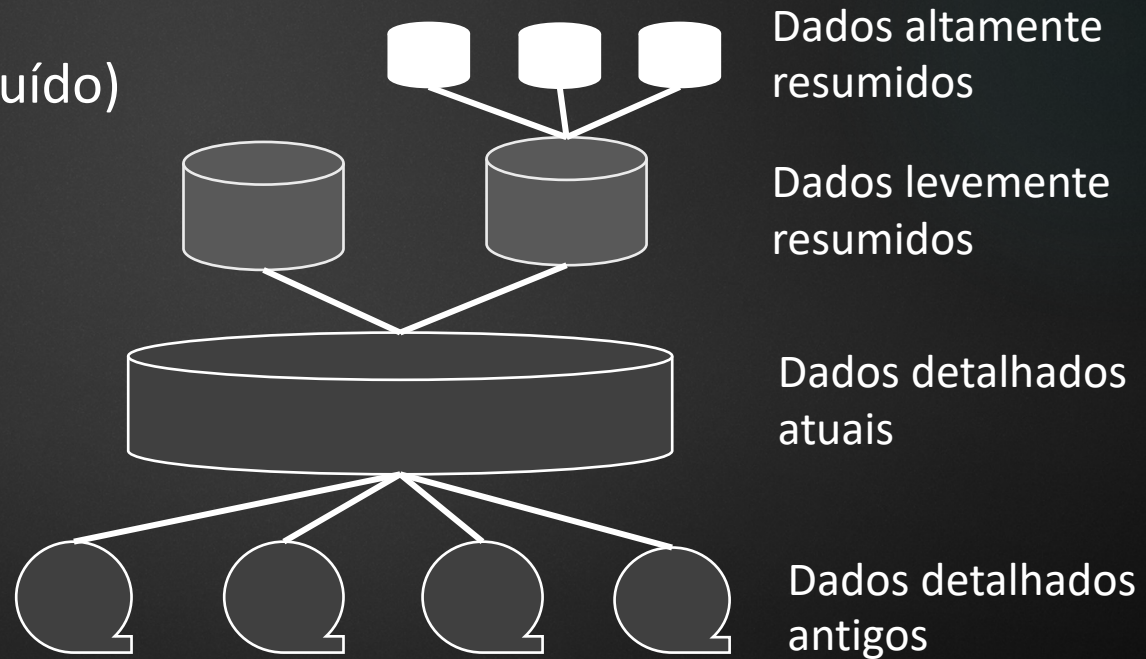
- Exemplo:

Baixa	Alta
01/01/2050 - R\$ 5000,00 02/01/2050 - R\$ 4500,00 03/01/2050 - R\$ 3850,00 31/01/2050 - R\$ 9800,00	01/2050: R\$ 120.000,00
01/02/2050 - R\$ 6850,00 02/02/2050 - R\$ 7500,00 28/02/2050 - R\$ 9500,00	02/2050: R\$ 112.650,00

2 - Data warehouse - DW

6-Localização

- Formas de armazenamento:
 - único local (centralizado)
 - por área de interesse (distribuído)
 - por nível de detalhes



2 - Data warehouse - DW


7-Credibilidade dos dados

- É o mais importante para o sucesso de qualquer projeto;
- Discrepâncias SIMPLES DE TODO TIPO podem causar sérios problemas quando se quer extrair dados para suportar decisões estratégicas para o negócio das empresas;
- Dados não dignos de confiança podem resultar em relatórios inúteis, que não tem importância alguma;
- Exemplo: Lista de pacientes do sexo MASCULINO e GRÁVIDOS;

2 - Data warehouse - DW

8-Metadados

- Para cada atributo do DW há uma entrada no dicionário de dados;
- Usuários ficam conhecendo a estrutura e o significado dos dados;
- No BD operacional, a estrutura e o significado dos dados estão embutidos nas aplicações;
- Repositórios de ferramentas CASE;



Qualidade dos dados

COMPÕE O DW

2 - Data warehouse - DW

Qualidade dos dados

- Baixa qualidade de dados é o maior obstáculo para o sucesso do gerenciamento do relacionamento com o cliente;
- Os problemas de qualidade de dados podem ser causados por dados redundantes e inconsistentes produzidos por múltiplos Sistemas;
- Erros de entrada de dados são a causa de muitos problemas de qualidade de dados;
- O DATA CLEANSING (limpeza e padronização) consiste em atividades para detectar e corrigir, no banco de dados, dados incorretos, incompletos, formatados inadequadamente ou redundantes.

Elementos

COMPÕE O DW

2 - Data warehouse - DW


Elementos

- Gerenciadores de Bancos de Dados
 - Relacionais: DB2, Oracle, SQL Server, Sybase, Informix, etc
 - Multidimensionais: Essbase, MS Olap Server, Oracle Express, etc
- Ferramentas de Modelagem (CASE)
 - ERWin, Paradigm, Rose, SA

2 - Data warehouse - DW

Elementos

- Ferramentas de Transformação e Carga
 - Pentaho, MS DTS, Ferramentas próprias, Extractor ETL, etc
- Ferramentas de Análise
 - Relatórios: através de linguagens como VB, Java, SQL, Pentaho, etc.
 - Queries: Oracle Express Analyzer, MS Office, Business Objects, Pentaho
 - DataMining



Datamart - DM

INTRODUÇÃO

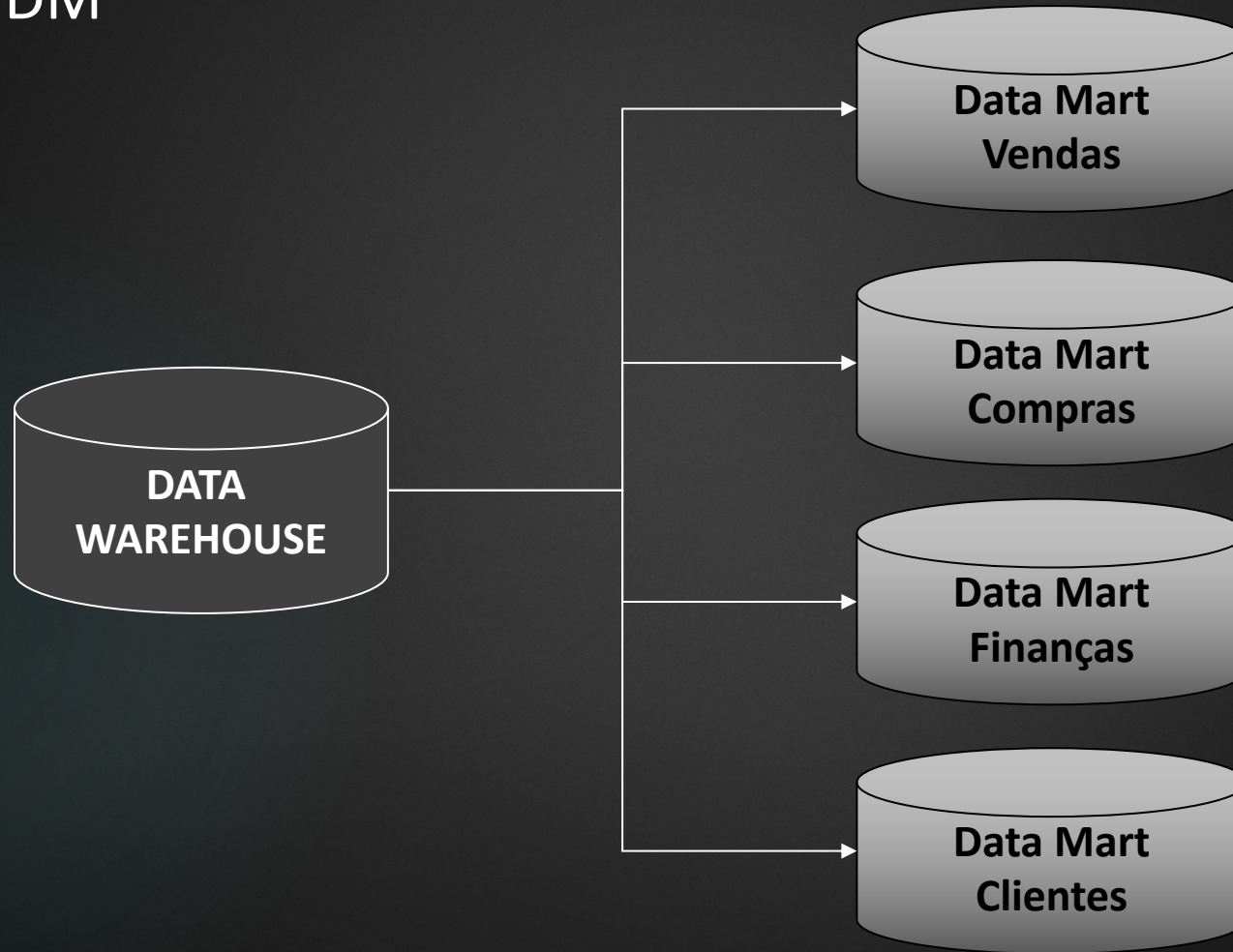
2 - Data warehouse - DW

Datamart - DM

- É um subconjunto de um DW;
- Representa um DW departamental, regional ou funcional;
- Em geral, são dados referentes a um assunto específico (ex: Vendas, Estoque, Compras, Clientes, Finanças, etc.).
- São bem aceitos no mercado por:
 - Apresentarem menor investimento em infra-estrutura;
 - Produzirem resultados mais rapidamente;
 - E por serem escaláveis até DWs;

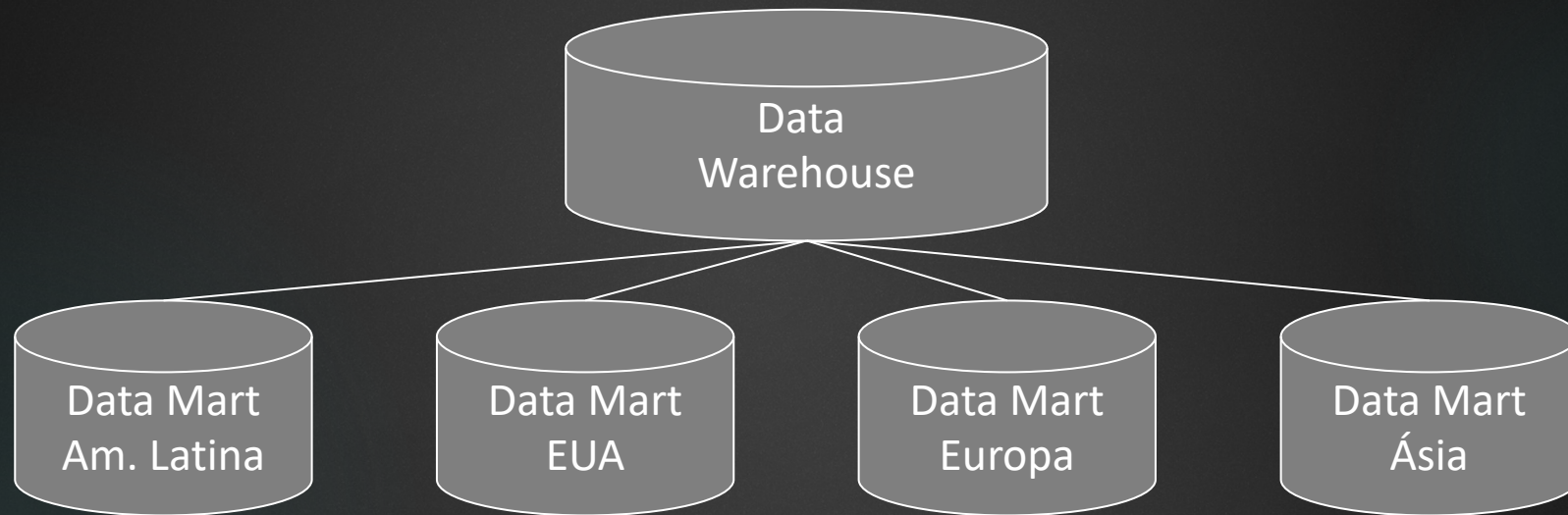
2 - Data warehouse - DW

Datamart - DM



2 - Data warehouse - DW

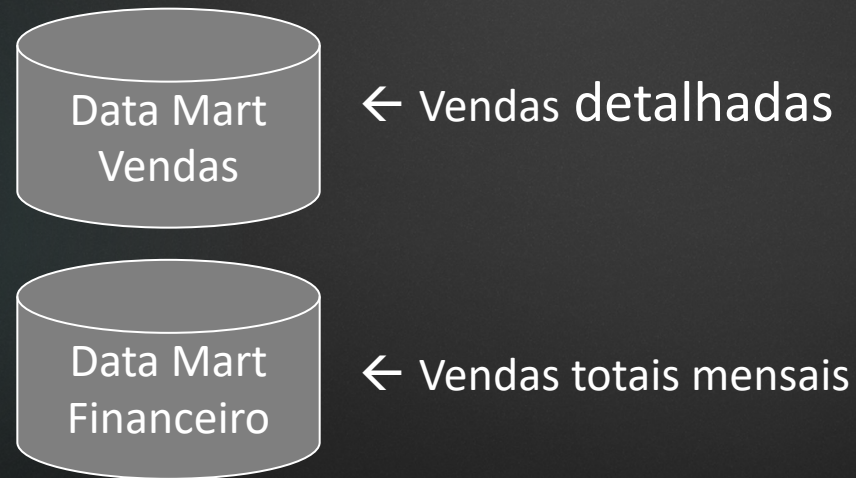
Datamart - DM



2 - Data warehouse - DW

Datamart - DM

- Dados podem ser repetidos em dois ou mais Data Marts;
- Os mesmos dados podem estar representados com granularidade diferente;
- Exemplo:





3- DATA WAREHOUSE x DATA LAKE

INTRODUÇÃO

3 – Data warehouse x Data Lake

Introdução

- Coletar, organizar, interpretar, criar insights e embasar decisões. São as essenciais para identificar oportunidades de melhorar os negócios;
- A base para todo esse processo é o armazenamento das informações e há duas formas para se fazer isso: Data Lake ou Data Warehouse;
- Essas tecnologias apresentam diferenças técnicas e conceituais em termos de arquitetura e finalidade.

2 – Data Lake (DL) x Data warehouse (DW)

Diferenças entre DL e DW – Formato dos dados

- Data Lake:
 - Armazenam todos os tipos de dados (estruturados, não estruturados e híbridos) em um só lugar;
 - Os dados iniciais são armazenados em seu formato bruto no data lake;
 - Depois, passam pelo processo ETL para carregar e transformar essas informações no data warehouse para análises futuras.
- Data warehouses:
 - São elementos obrigatórios em soluções de Big Data de grande escala;
 - Armazenam dados estruturados

2 – Data Lake (DL) x Data warehouse (DW)

Diferenças entre DL e DW – Armazenamento

- Data Lake:
 - São repositórios enormes, capazes de guardar grandes quantidades de dados estruturados e não estruturados.
 - Armazenam os dados brutos: sem verificações, processamentos, governanças ou análises
- Data warehouses:
 - Contêm apenas dados processados, higienizados e verificados.

2 – Data Lake (DL) x Data warehouse (DW)

Diferenças entre DL e DW – Usuários

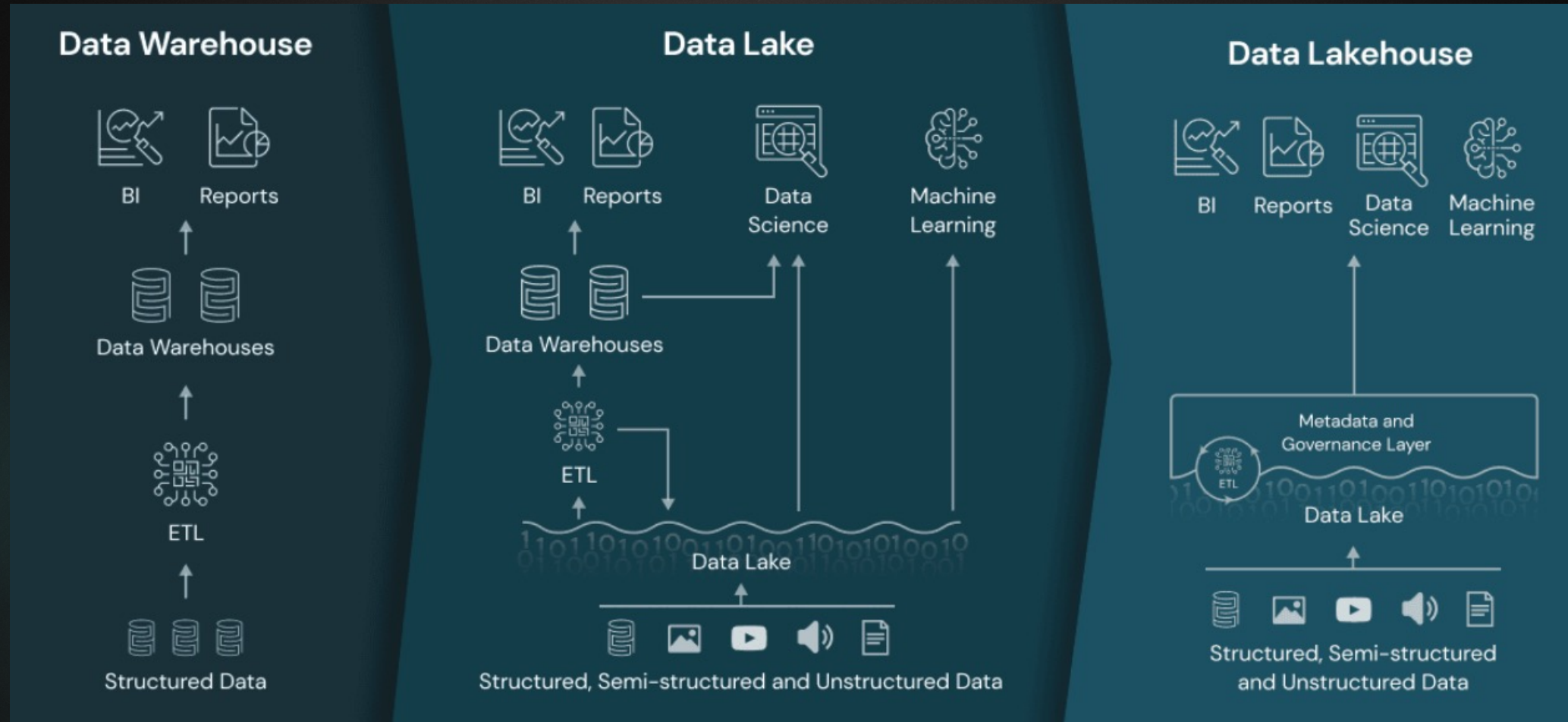
- Data Lake:
 - Engenheiros e cientistas de dados;
- Data warehouses:
 - Analistas de negócios e stakeholders;

2 – Data Lake (DL) x Data warehouse (DW)

Diferenças entre DL e DW

- Data Lake tem dados brutos armazenados sem um fim, já o Data Warehouse tem dados estruturados para finalidades preexistentes;
- Data Lake é um repositório para manter informações dos mais diferentes tipos que, no futuro, serão úteis para captar ideias necessárias na empresa;
- Data Warehouse é um recurso para solucionar problemas e encontrar alternativas no momento de decisão;

2 – Data Lake (DL) x Data warehouse (DW)





Obrigado!