

AULA 2

- **Dados | Informação | Conhecimento**
- **Big Data**

Dados

- **Definição:** Dados são fatos brutos e objetivos, geralmente representados por números, letras ou símbolos, que ainda não foram processados ou interpretados.
- **Exemplo:** Números de vendas mensais, registros de transações, ou qualquer informação crua sem contexto.

Informação

- **Definição:** Informação é o resultado do processamento e organização dos dados, dando-lhes contexto e significado.
- **Exemplo:** Relatório de vendas mensais que destaca as tendências ao longo do tempo ou gráficos que mostram a distribuição geográfica dos clientes.

Conhecimento

- **Definição:** Conhecimento é uma compreensão mais profunda e interpretativa da informação, associada à experiência e contextualização.
- **Exemplo:** Análise de mercado que identifica padrões sazonais nas vendas, insights sobre o comportamento do cliente e as razões por trás das flutuações.

Inteligência

- **Definição:** Inteligência é a capacidade de utilizar o conhecimento para tomar decisões informadas e antecipar eventos futuros.

- **Exemplo:** Implementação de estratégias de marketing baseadas nas análises para maximizar as vendas em períodos específicos ou a previsão de demanda para otimizar o estoque.

Inteligência de Negócios (Business Intelligence – BI)

Inteligência de Negócios, também conhecida como Business Intelligence (BI), refere-se ao conjunto de processos, tecnologias e ferramentas que ajudam as organizações a coletar, analisar e transformar dados brutos em informações significativas para suportar tomadas de decisões empresariais.

Os principais objetivos da Inteligência de Negócios são fornecer insights acionáveis, melhorar o processo decisório e impulsionar o desempenho organizacional. Isso é alcançado por meio de várias atividades, como:

1. **Coleta de Dados:** Reúne dados relevantes de diversas fontes, tanto internas quanto externas à organização.
2. **Processamento e Transformação de Dados:** Organiza e prepara os dados para análise, incluindo a limpeza, integração e estruturação adequada.
3. **Análise de Dados:** Utiliza técnicas estatísticas e ferramentas analíticas para identificar padrões, tendências e insights nos dados.
4. **Visualização de Dados:** Apresenta os resultados de forma clara e compreensível por meio de gráficos, dashboards e relatórios interativos.
5. **Tomada de Decisões Informadas:** Fornece informações relevantes e oportunas para apoiar as decisões estratégicas e operacionais da organização.

As soluções de BI podem ser implementadas por meio de diferentes ferramentas e plataformas especializadas, muitas das quais oferecem funcionalidades avançadas, como análise preditiva e mineração de dados. O objetivo final é capacitar os usuários, desde executivos até analistas de dados, a entenderem melhor o desempenho da empresa, identificarem oportunidades e anteciparem desafios.

A Inteligência de Negócios desempenha um papel crucial em ambientes empresariais modernos, ajudando as organizações a aproveitar ao máximo seus dados para obter vantagens competitivas e melhorar continuamente suas operações.

BIG DATA

Big data refere-se a conjuntos de dados **extremamente grandes e complexos** que não podem ser facilmente processados, gerenciados ou analisados com ferramentas de processamento de dados tradicionais. Esses conjuntos de dados possuem **características específicas** conhecidas como os "**Três V's**": **Volume, Velocidade e Variedade**.

- **Volume:** Refere-se à imensa quantidade de dados gerados continuamente. Por exemplo, as redes sociais, transações financeiras online, dispositivos IoT (Internet das Coisas) e dados de sensores industriais geram volumes massivos de informações a cada segundo.
- **Velocidade:** Diz respeito à taxa rápida na qual os dados são gerados, processados e disponibilizados para análise. Um exemplo é o fluxo constante de dados em tempo real gerado por transações online, feeds de redes sociais e dispositivos conectados.
- **Variedade:** Indica a diversidade de fontes e tipos de dados. Além dos dados estruturados presentes em bancos de dados convencionais, o big data engloba dados não estruturados, como textos, áudio e vídeo, e dados semi-estruturados, como JSON e XML.

Exemplos de aplicação de *big data*

Análise de Redes Sociais: Plataformas como Facebook, Twitter e Instagram geram grandes volumes de dados sobre interações sociais, preferências e comportamentos dos usuários.

Análise Financeira: Instituições financeiras utilizam big data para analisar transações em tempo real, detectar padrões de fraude e realizar avaliações de risco de maneira mais precisa.

Saúde: Hospitais e organizações de saúde aplicam big data para analisar registros médicos eletrônicos, resultados de testes e dados genômicos, facilitando diagnósticos mais rápidos e personalizados.

Manufatura: Empresas podem utilizar big data para monitorar e otimizar processos de produção, analisando dados de sensores e equipamentos em tempo real.

Comércio Eletrônico: Plataformas de comércio eletrônico utilizam big data para recomendações personalizadas com base no histórico de compras e comportamento do usuário.

O *big data*, quando gerenciado e analisado adequadamente, pode fornecer *insights* valiosos, impulsionar a inovação e apoiar decisões estratégicas em diversos setores.

Dados estruturados vs *Big data*

A diferença fundamental entre **big data** e **dados estruturados** reside na natureza dos dados e na forma como são organizados, armazenados e processados. Veja:

Big Data

- **Natureza dos Dados:** Big data abrange conjuntos de dados extremamente grandes e complexos, caracterizados pelos "Três V's" (Volume, Velocidade e Variedade).
- **Volume:** Envolvem grandes quantidades de dados, muitas vezes na ordem de pentabytes ou exabytes.
- **Velocidade:** São gerados em alta velocidade, exigindo processamento em tempo real para análise.
- **Variedade:** Podem incluir dados estruturados, não estruturados e semi-estruturados de diversas fontes.

Dados Estruturados

- **Natureza dos Dados:** Dados estruturados são altamente organizados e seguem um modelo predefinido, geralmente baseado em tabelas relacionais¹.
- **Formato:** São organizados em campos, registros e tabelas, com relações claramente definidas.
- **Exemplos:** Bancos de dados relacionais são uma forma comum de dados estruturados. Por exemplo, uma tabela em um banco de dados SQL que armazena informações sobre clientes, onde cada linha representa um cliente e as colunas representam atributos como nome, idade, e e-mail.

¹ Em um **banco de dados relacional**, as informações são organizadas em **tabelas**, onde cada tabela consiste em linhas e colunas. O termo "**relacional**" refere-se à **capacidade de estabelecer relações entre diferentes tabelas**.

Exemplo Comparado

Vamos considerar um cenário em que uma empresa de comércio eletrônico está coletando dados sobre transações de clientes:

Dados Estruturados	Big Data
As informações específicas sobre cada transação, como data, produto comprado, preço, e número de identificação do cliente, seriam exemplos de dados estruturados. Esses dados seriam armazenados em um banco de dados relacional de forma organizada e predefinida.	Nesse contexto, o <i>big data</i> incluiria não apenas as transações estruturadas (como informações em tabelas de banco de dados), mas também dados não estruturados, como feedbacks em redes sociais, análises de sentimentos dos clientes e até mesmo dados de streaming em tempo real durante eventos de vendas.

Em resumo, enquanto **dados estruturados têm uma organização rígida e previsível**, o **big data envolve conjuntos massivos e diversos**, abrangendo uma gama mais ampla de tipos de dados. O *big data* é mais desafiador de ser processado devido à sua complexidade e ao seu tamanho, exigindo abordagens e ferramentas específicas para extrair informações úteis.

Processo de Tratamento de Dados

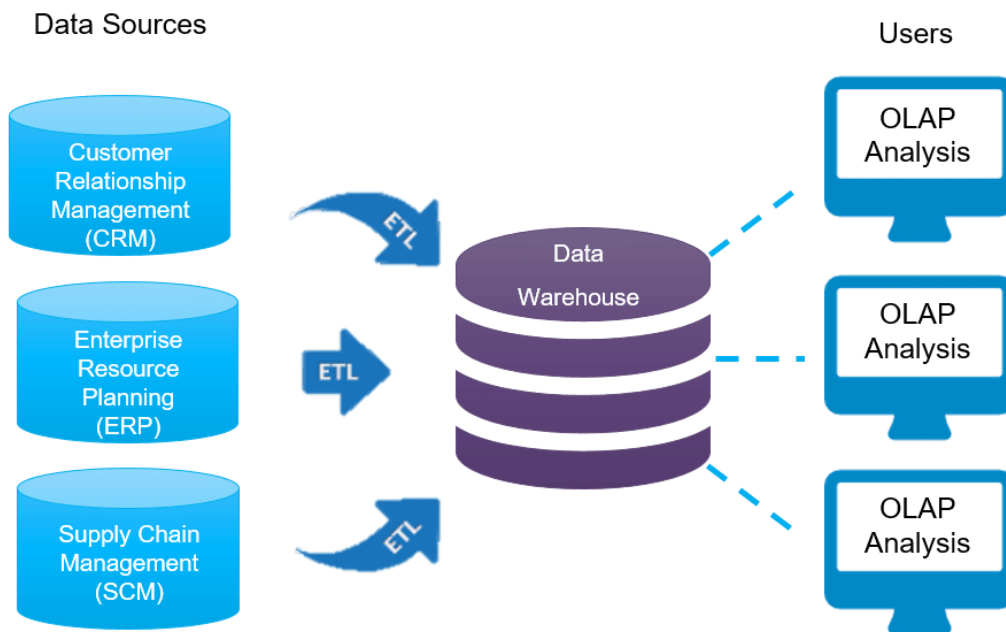
O tratamento de **big data** envolve várias etapas, desde a coleta até a análise, e inclui o **armazenamento** eficiente desses enormes conjuntos de dados.

Os conceitos de **Data Lake** e **Data Warehouse** são peças importantes nesse processo, cada um desempenhando funções distintas.

Um **Data Lake** é mais flexível e **armazena** dados brutos em seu formato original, enquanto um **Data Warehouse** é otimizado para análises estruturadas e consultas eficientes, **fornece uma visão consolidada dos dados para suportar decisões de negócios**. Ambos os conceitos podem coexistir em uma arquitetura de dados abrangente, cada um desempenhando um papel específico no ecossistema de gerenciamento de dados da organização.

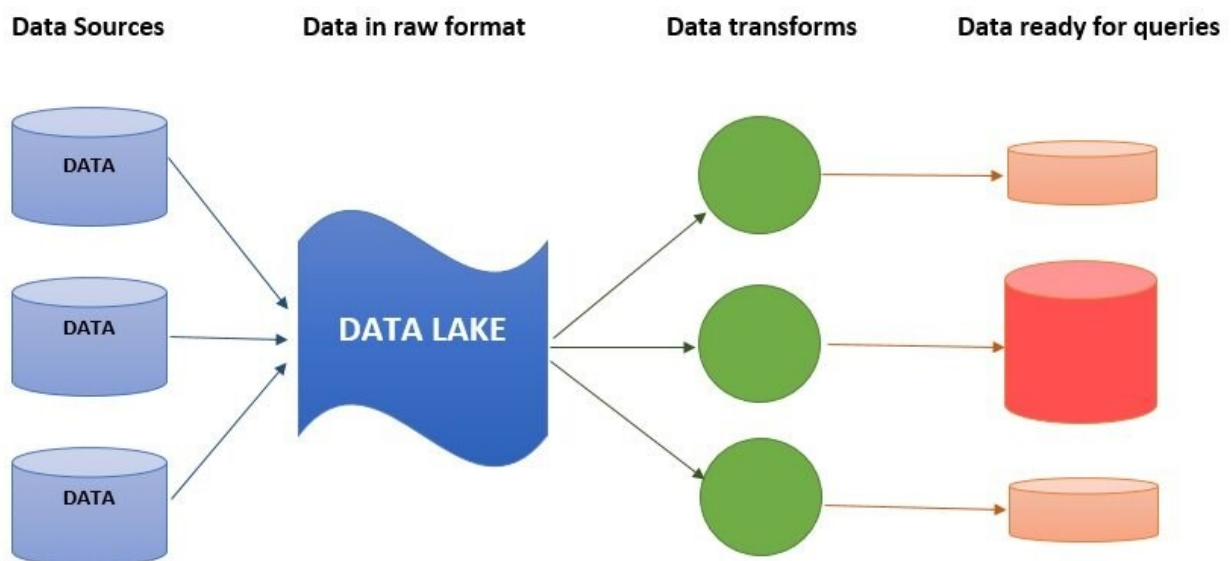
Sugestão → Faça a leitura no site da Amazon sobre Data Lake: <https://aws.amazon.com/pt/blogs/aws-brasil/como-extrair-valor-dos-seus-dados-com-data-lakes-e-analises-na-aws/>

DATA WAREHOUSE



<https://intellipaas.com/blog/tutorial/data-warehouse-tutorial/data-warehouse-overview/>

DATA LAKE



Fonte: Databricks

Data Lake	Data Warehouse
<p>Repositório centralizado e expansível que armazena dados brutos em seu formato original, independentemente de sua origem ou estrutura.</p> <p>Projetado para lidar com grandes volumes de dados, incluindo dados estruturados, não estruturados e semi-estruturados.</p> <p>Pode armazenar uma variedade de dados, desde logs de servidores e transações até documentos de texto não estruturados.</p>	<p>Sistema de gerenciamento de dados que armazena dados organizados e estruturados, otimizados para análises e consultas eficientes. Os dados são transformados e agregados para fornecer uma visão coesa e consistente dos dados para suportar a tomada de decisões.</p> <p>Data Warehouse é utilizado para análises estruturadas e relatórios de negócios.</p>
<p>Exemplo</p> <p>Uma empresa de comércio eletrônico pode usar um Data Lake para armazenar registros brutos de interações do cliente, logs de servidores, dados de transações e até mesmo feedbacks não estruturados dos clientes, todos em seu formato original. Isso permite que a empresa explore e analise esses dados de maneira flexível para obter insights mais profundos sobre o comportamento do cliente, padrões de compra e tendências de mercado.</p>	<p>Exemplo</p> <p>Considerando a mesma empresa de comércio eletrônico, o Data Warehouse poderia armazenar dados consolidados, como vendas mensais por região, segmentação de clientes com base no histórico de compras, e métricas de desempenho de marketing. Esses dados são transformados e organizados de maneira a facilitar consultas analíticas, relatórios gerenciais e análises de tendências de longo prazo.</p>

Podemos ainda falar em **Data Marts**, que são subconjuntos especializados de um *data warehouse* maior, focados em atender às necessidades específicas de um grupo particular de usuários ou de um departamento dentro de uma organização. Enquanto um *data warehouse* abrange uma variedade de dados para suportar análises corporativas amplas, os **data marts** são projetados para fornecer informações mais específicas e detalhadas a um público-alvo mais restrito.

Principais características dos **Data marts**:

Escopo Específico: Cada *data mart* é direcionado a atender as demandas de um setor ou grupo de usuários específico, como vendas, marketing, finanças, etc.

Dados Especializados: Contêm dados relevantes e específicos para o domínio ou departamento para os quais foram criados. Esses dados são extraídos do *data warehouse* ou, em alguns casos, diretamente das fontes originais.

Facilidade de Acesso: São estruturados para facilitar o acesso e a análise por parte dos usuários destinados. Isso permite que esses usuários encontrem mais rapidamente as informações relevantes para suas atividades.

Autonomia: Os data marts podem ser desenvolvidos de forma independente, o que oferece mais agilidade na resposta às necessidades específicas de um departamento sem afetar o data warehouse central.

Desempenho Otimizado: Como são mais especializados e menores em escopo, os data marts podem ser otimizados para proporcionar consultas mais rápidas e eficientes aos usuários finais.

Flexibilidade: Permitem adaptações específicas para atender aos requisitos particulares de um departamento, sem interferir nos outros segmentos da organização.

Existem dois principais tipos de data marts:

Dependent Data Mart: Esses data marts são alimentados diretamente pelo data warehouse central. São mais adequados para ambientes onde a estrutura de dados é estável e controlada centralmente.

Independent Data Mart: Desenvolvidos de forma autônoma, esses data marts extraem dados diretamente das fontes operacionais. São mais adequados quando há necessidade de agilidade e autonomia em ambientes descentralizados.

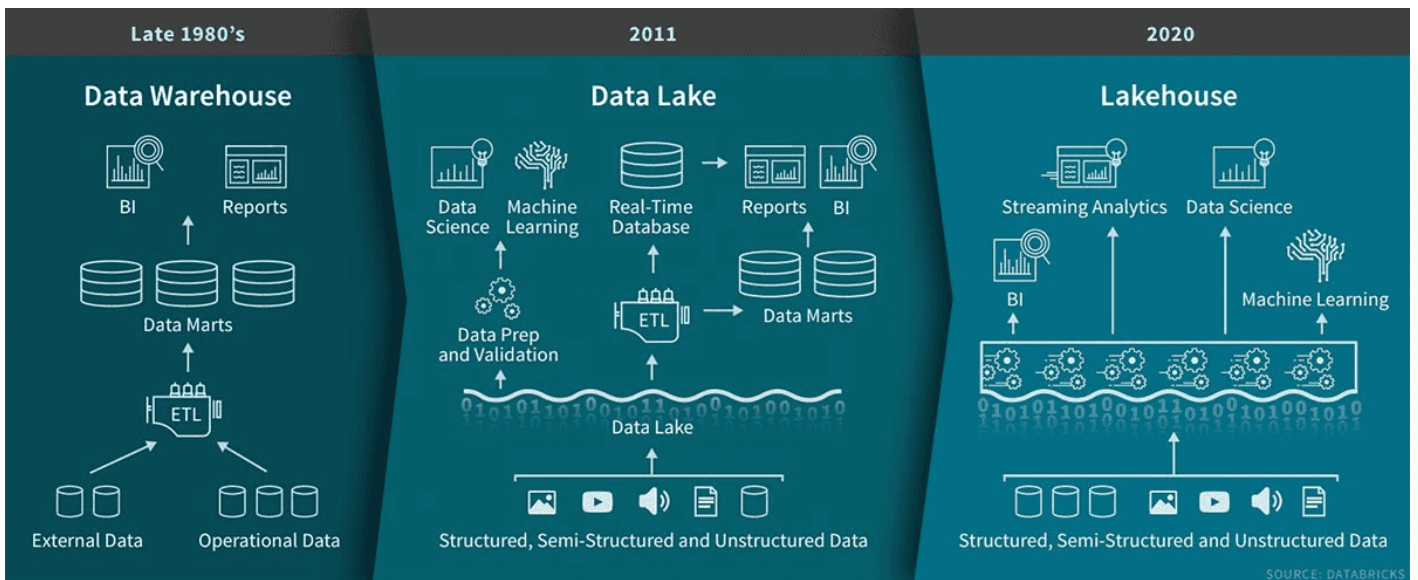
Os *data marts* são uma estratégia eficaz para fornecer informações personalizadas e orientadas para a ação a áreas específicas de uma organização, contribuindo para a tomada de decisões mais informada e eficiente.

ADEQUAÇÃO DAS ESTRUTURAS ÀS NECESSIDADES EM CONSTANTE MUDANÇA



Definição de um Data Lakehouse. Fonte: Oracle.

Evolução → Data Lakehouse



Fonte: Databricks

REFERÊNCIAS

FERREIRA, Rafael Gastão Coimbra... [et al.] Preparação e análise exploratória de dados; revisão técnica: Jean Paul Barddal. Porto Alegre: SAGAH, 2021.

LAUDON, K. C., & LAUDON, J. P. . Sistemas de Informação Gerenciais. São Paulo: Pearson, 2016.