

CONHECENDO OS DADOS

COMO OS DADOS SE TRANSFORMAM
EM INFORMAÇÃO E DECISÕES



Valdiney França

VALDINEY FRANÇA

CONHECENDO OS DADOS

Como os dados se transformam
em informação e decisões

Copyright

© 2026 Valdiney França

Todos os direitos reservados.

Nenhuma parte desta publicação poderá ser reproduzida, distribuída ou transmitida, por qualquer meio, sem autorização prévia e por escrito do autor, exceto nos casos previstos em lei.

1ª edição – 2026

Autor: Valdiney França

Título: Conhecendo os Dados

ISBN: 978-65-02-20361-3

Revisão

Adriana Nolasco: Gerente de Projetos de TI

Leonardo Fernandez: Engenheiro de Dados

Sobre o Autor

Valdiney França é desenvolvedor de software e atua há mais de uma década na criação, manutenção e evolução de sistemas corporativos. Ao longo da carreira, participou de projetos nas áreas de educação, saúde, ERP, PDV, CRM e Internet das Coisas (IoT), acompanhando desde a análise de requisitos até a implementação e modernização de aplicações em produção.

Seu trabalho passa pela organização de sistemas, modelagem de dados e refatoração de aplicações legadas, sempre com foco em soluções mais simples de manter, evoluir e compreender. Ao longo dos anos, participou de

projetos de diferentes portes e contribuiu para sistemas usados por milhares de pessoas.

Mais do que desenvolver funcionalidades, acredita que o papel do software é organizar informações, apoiar decisões e gerar valor para pessoas e organizações. Para ele, tecnologia não se resume a código. Sistemas bem estruturados ajudam a transformar conhecimento em ação.

Esta obra nasce da experiência prática em projetos reais e da convicção de que compreender dados, processos e regras de negócio é um dos primeiros passos para criar soluções mais eficientes e sustentáveis.

Contato

Valdiney França

E-mail: valdineyweb@gmail.com

LinkedIn: [linkedin.com/in/valdiney-frança](https://www.linkedin.com/in/valdiney-frança)



Sumário

Sobre o Autor.....	4
Introdução.....	8
Importância dos Dados.....	10
Dados em Informação.....	15
Informação, Ruído e Interpretação.....	22
Fenômeno Big Data.....	28
Estrutura dos Dados.....	35
O que é um Banco de Dados.....	45
Ciclo de Vida dos Dados.....	50
Análise de Dados.....	56
Quem cuida dos Dados?.....	59
Qualidade e Ética.....	70
LGPD.....	74
Considerações Finais.....	88
Referências.....	91

Introdução

Você esbarra em dados o tempo todo. Eles aparecem no aplicativo que sugere a próxima música, no GPS que recalcula a rota, na loja que registra uma venda e até nas decisões de uma empresa ao contratar ou demitir alguém. Também aparecem na gestão pública, quando governos analisam números para definir políticas e investimentos.

Os dados também estão na base da Inteligência Artificial. Sistemas de IA aprendem padrões a partir de grandes volumes de dados e usam esses padrões para reconhecer imagens, sugerir textos, responder perguntas ou apoiar decisões. Por isso, quando falamos sobre dados,

também estamos falando sobre a matéria-prima que torna muitas soluções de IA possíveis.

Mesmo usando dados todos os dias, nem sempre paramos para entender o que eles são, como viram informação e por que influenciam tantas decisões.

Este livro parte desse básico. A ideia não é ensinar uma ferramenta específica, mas construir uma base: entender o que é um dado, quando ele vira informação e por que esse caminho influencia sistemas, empresas e decisões.

A partir daí, entramos em temas como importância dos dados, diferença entre dado e informação, interpretação, Big Data, estruturas de dados, bancos de dados, análise, qualidade e LGPD

Importância dos Dados

Dados sempre tiveram valor para pessoas, governos e empresas.

Desde os primeiros registros da história, o ser humano busca guardar informações e, muitas vezes, controlar quem pode acessá-las. Isso não acontecia por acaso. Quem detém dados detém conhecimento, e quem detém conhecimento costuma ocupar posições estratégicas no mercado e na sociedade.

Controlar a informação sempre foi uma forma de exercer poder, tomar decisões melhores e antecipar movimentos. Em alguns

momentos da história, essa vantagem foi tão grande que a própria informação, verdadeira ou falsa, passou a ser usada como instrumento direto de manipulação.

Ao longo da história, houve momentos em que informações falsas não apenas influenciaram decisões, mas foram usadas de propósito para gerar lucro ou vantagem. Um exemplo menos conhecido ocorreu em 1814, no episódio que ficou conhecido como Great Stock Exchange Fraud of 1814. Na ocasião, notícias fabricadas sobre a morte de Napoleão Bonaparte e a suposta derrota da França foram espalhadas em Londres. Ao acreditarem nisso, comerciantes e investidores mudaram rapidamente seu comportamento financeiro. Enquanto a verdade ainda não havia sido verificada, aqueles que criaram e controlaram a

informação obtiveram ganhos substanciais antes que a realidade viesse à tona. O episódio mostra que o valor não está apenas na informação em si, mas no tempo, no controle e no uso que se faz dela.

Se, em 1814, uma notícia falsa levava dias para produzir efeitos relevantes no mercado financeiro londrino, na era digital algo parecido pode acontecer em segundos e em escala global. Informações circulam instantaneamente, são amplificadas por plataformas digitais e podem influenciar decisões automatizadas em mercados, sistemas de crédito e cadeias produtivas. A diferença não está só no conteúdo da informação, mas na velocidade, no alcance e na automação dos sistemas que reagem a ela.

Por isso, armazenar, processar e proteger dados não é detalhe técnico. Dados confiáveis podem virar informação; informação bem usada sustenta decisões. Sem dados, não há informação. Sem informação confiável, sistemas computacionais perdem parte importante de sua função: apoiar a tomada de decisão.

Mas há uma limitação importante: dados, sozinhos, não garantem boas decisões. Dados incompletos, desatualizados ou incorretos podem levar a conclusões erradas, mesmo quando passam por sistemas sofisticados. Ter dados não é o mesmo que compreender a realidade. É só o começo. Com a digitalização dos processos, cada clique, transação ou interação produz novos registros. Sistemas modernos não apenas armazenam informações;

eles dependem delas para operar em tempo real, automatizar decisões e antecipar cenários. Nesse ponto, entender dados deixa de ser apenas uma questão técnica.

Para entender esse valor, precisamos começar pelo básico: o que é um dado e em que momento ele se transforma em informação.

Dados em Informação

Você provavelmente já ouviu as palavras dados e informação em vários contextos. Muita gente usa as duas como se fossem a mesma coisa, mas não são. Essa diferença parece simples, porém muda a forma como enxergamos sistemas, relatórios e decisões.

Um dado é o registro bruto de um fato. Ele representa algo que foi observado, medido ou coletado, mas ainda não foi interpretado. Pode ser um número, uma data, um texto curto ou qualquer valor armazenado por um sistema.

A informação surge quando um ou mais dados são organizados e colocados em um

contexto que permita compreensão. A informação comunica algo. Ela responde a uma pergunta e reduz incertezas.

Considere a letra “S”. Isolada, ela é apenas um dado: sabemos que é uma letra do alfabeto, mas não há um significado específico associado a ela. Ao juntarmos outras letras, S-A-L-V-A-D-O-R, formamos uma palavra. Ainda assim, sem contexto, essa palavra pode representar coisas diferentes.

Quando afirmamos que Salvador foi a primeira capital do Brasil, aquelas letras deixam de ser apenas símbolos organizados e passam a comunicar um fato histórico. Nesse momento, os dados foram organizados, contextualizados e interpretados. Tornaram-se informações.

Em termos simples, dados podem ser entendidos como matéria-prima, enquanto informação é o resultado do processamento dessa matéria-prima. Um dado sem contexto é como uma palavra fora de uma frase: ele existe, mas não comunica plenamente. A informação é o conjunto estruturado que permite entendimento.

Esse mesmo raciocínio aparece na linguagem. Imagine uma pessoa falante da língua portuguesa tentando ler um livro em árabe. Para ela, aquelas letras são apenas símbolos sem significado. Mesmo organizadas em palavras e frases, continuam sendo dados, porque não existe compreensão. Para um falante da língua árabe, porém, aqueles mesmos símbolos seguem regras gramaticais conhecidas, formam palavras, frases e

transmitem ideias. Nesse caso, os dados são interpretados e se transformam em informação.

O mesmo princípio se aplica a sistemas empresariais. Imagine um banco de dados que armazena o número 1500. Isoladamente, esse valor é apenas um dado. Ele não comunica nada além de um número registrado.

Quando esse número é associado ao campo “valor de venda” de um determinado produto, passa a representar uma transação realizada. Nesse contexto, o dado ganha significado e se transforma em informação.

Se o sistema identifica que esse valor corresponde ao total vendido de um produto específico no mês, e que no mês anterior o total foi 1000, a comparação revela crescimento. A partir daí, os dados organizados ajudam a

compreender o desempenho do produto e a apoiar decisões, como aumentar o estoque ou investir em marketing.

Em todos esses exemplos, o conteúdo original não mudou. O que mudou foi o contexto e a possibilidade de interpretação. Dados se transformam em informação quando são lidos dentro de uma estrutura que lhes dá significado.

Nem toda interpretação, porém, produz informação confiável. Interpretações equivocadas podem gerar conclusões estruturadas, mas não necessariamente verdadeiras. A existência de contexto não garante correção; ela apenas permite significado. A qualidade da informação depende da qualidade dos dados e da forma como eles são analisados.

Essa diferença não serve apenas para uma definição de livro. Ela ajuda a entender como registros brutos podem caminhar até decisões e ações concretas. Um dos modelos mais usados para explicar essa progressão é a Hierarquia DIKW, que organiza o processo em quatro níveis: Dados, Informação, Conhecimento e Sabedoria. O modelo ajuda, mas não deve ser tratado como verdade absoluta.

Nesse modelo, os dados são registros brutos. A informação surge quando esses dados recebem contexto e interpretação. O conhecimento aparece quando a informação é aplicada de maneira consistente em situações práticas. A sabedoria envolve a capacidade de tomar decisões adequadas a partir desse conhecimento, considerando experiência, julgamento e um contexto mais amplo.

Quando essa diferença não está clara, um sistema pode guardar milhares de registros e ainda assim dizer muito pouco. Quantidade não substitui entendimento. Esse cuidado evita um erro comum: achar que a tecnologia, sozinha, resolve problemas de análise e decisão.

Informação, Ruído e Interpretação

Informação não é o fato em si. Ela nasce da interpretação de dados que representam fatos.

Os fatos existem independentemente de opinião. São acontecimentos ou condições do mundo real. Quando um sistema registra uma venda, quando um sensor mede uma temperatura ou quando um cliente realiza um pagamento, algo ocorreu. O dado nasce como registro desse ocorrido. Ele é a marca formal de que um fato aconteceu.

A informação surge em um passo posterior. Ela aparece quando esses dados são organizados, contextualizados e interpretados. Um número isolado diz pouco. Mas quando sabemos o que ele representa, a que período pertence e com o que pode ser comparado, ele começa a ganhar significado.

Essa discussão se aproxima da Teoria da Informação, mas não é a mesma coisa. A teoria investiga como sinais são transmitidos, armazenados e recebidos de forma eficiente. O foco está na integridade da transmissão e na redução da incerteza técnica, não na interpretação humana.

Claude Shannon, fundador da Teoria da Informação, demonstrou que a informação está presente sempre que um sinal é transmitido de

um emissor para um receptor, independentemente de seu significado. Para esse modelo, o conteúdo semântico é irrelevante. O que importa é a estrutura do sinal e sua probabilidade de ocorrência.

Quando tratamos de informação no contexto humano ou computacional voltado à tomada de decisão, o significado passa a ser central. Um sistema pode transmitir dados com precisão absoluta, sem perda de pacotes, sem falhas no armazenamento e sem corrupção de bits, e ainda assim não gerar informação útil para quem recebe esses dados.

Entre o dado transmitido e a informação interpretada existe um elemento decisivo: o ruído.

Ruído é qualquer fator que interfira na comunicação ou distorça a compreensão. Ele pode ser técnico, como falhas no canal de transmissão, interferências ou perdas de sinal. Também pode ser interpretativo: ambiguidades, ausência de contexto, diferenças culturais, pressupostos incorretos ou limitações cognitivas.

Mesmo quando não há falha técnica, o ruído interpretativo pode impedir que os dados se transformem em informação. Um relatório pode estar matematicamente correto e, ainda assim, ser mal compreendido. Nesse caso, o problema não está na transmissão, mas na interpretação.

A informação pode ser transmitida de várias formas. Uma música é uma sequência

organizada de sinais sonoros que, ao serem percebidos, produzem sentido. Palavras são símbolos organizados segundo regras linguísticas, capazes de construir significados complexos. Em sistemas computacionais, sinais assumem a forma de bits organizados em estruturas formais.

Em todos esses casos, há algo em comum: sinais são emitidos, recebidos e, muitas vezes, registrados. Isso significa que a informação não apenas circula; ela também pode ser armazenada, preservada e reutilizada. Os sistemas digitais ampliaram essa capacidade de registro em uma escala inédita.

Quando a capacidade de registrar e armazenar cresce, outro problema aparece. Não basta transmitir bem um dado. É preciso lidar

com volume, velocidade, formatos diferentes e ruídos que também crescem junto com a escala.

É nesse cenário que surge o Big Data.

Fenômeno *Big Data*

A produção de dados não começou com computadores. Pinturas rupestres, inscrições em pedra, papiros e livros contábeis já eram formas de registrar fatos e preservar informações. O hábito é antigo; a diferença está no tamanho que ele ganhou.

O que mudou não foi o hábito de registrar, mas a escala.

Hoje, quase toda atividade deixa rastros digitais. Redes sociais, transações financeiras, partidas esportivas, sensores industriais, celulares e sistemas corporativos geram dados o tempo inteiro. A produção deixou de acontecer de vez em quando e passou a ser contínua. O

volume cresceu tanto que muitos métodos tradicionais de armazenamento e processamento já não dão conta.

O termo Big Data começou a ganhar visibilidade em 1998, quando John Mashey, então cientista-chefe da Silicon Graphics, passou a utilizá-lo para descrever os desafios associados ao grande volume de dados. A partir dos anos 2000, com a expansão da internet, dos sistemas distribuídos e dos dispositivos conectados, o fenômeno se tornou mais evidente. Por volta de 2010, consolidou-se como conceito de mercado, associado a novas arquiteturas, plataformas e estratégias orientadas a dados.

Inicialmente, *Big Data* foi descrito a partir dos chamados 3 V's:

Volume: a quantidade de dados produzidos é massiva. Não se trata mais de milhares de registros, mas de milhões ou bilhões. Cada clique, cada transação e cada evento adiciona novos dados ao sistema.

Velocidade: os dados são gerados e precisam ser processados rapidamente. Em muitos contextos, a análise deve ocorrer quase em tempo real, como na detecção de fraudes ou no monitoramento de sistemas críticos.

Variedade: os dados assumem múltiplos formatos. Além de tabelas estruturadas, há textos livres, imagens, vídeos, áudios, registros de log e sinais de sensores. Nem tudo pode ser organizado facilmente em linhas e colunas.

Com o amadurecimento do uso de dados em larga escala, percebeu-se que essas três

dimensões não bastavam para descrever o fenômeno por completo. Dois novos elementos passaram a aparecer com frequência no modelo:

Veracidade: refere-se à qualidade e confiabilidade dos dados. Em ambientes de grande escala, erros, inconsistências, duplicações e vieses podem se propagar rapidamente, comprometendo análises e decisões.

Valor: diz respeito à capacidade de transformar dados em benefícios concretos. Grandes volumes só justificam investimento quando resultam em melhorias operacionais, vantagem competitiva ou geração de conhecimento relevante.

Com isso, o modelo passou de três para cinco dimensões. Ele deixou de olhar só para o

desafio técnico e passou a considerar também qualidade e finalidade.

O desafio do Big Data não está apenas em armazenar grandes volumes, mas em processá-los, analisá-los e extrair informação relevante dentro de um tempo útil. À medida que volume, velocidade e variedade aumentam, cresce também a complexidade. Para lidar com heterogeneidade e escala, são necessárias novas arquiteturas, processamento distribuído, métodos estatísticos e algoritmos adequados.

Mas mais dados não significam automaticamente mais informação.

Como vimos antes, a informação depende de interpretação e contexto. Em ambientes de Big Data, o risco de ruído também cresce. Correlações enganosas, padrões irrelevantes e

interpretações apressadas podem gerar conclusões equivocadas, mesmo quando os dados foram coletados corretamente.

Acumular dados é relativamente simples. Transformá-los em informação confiável exige capacidade técnica, método analítico e clareza sobre o problema que se deseja resolver.

A Inteligência Artificial tornou essa discussão ainda mais visível. Muitos modelos precisam de grandes volumes de dados para serem treinados, avaliados e melhorados. Mesmo assim, o princípio continua o mesmo: quantidade só gera valor quando existe qualidade, contexto e interpretação.

Antes de aplicar técnicas de análise ou mineração, é preciso olhar para a natureza dos dados disponíveis. Alguns são altamente

estruturados e ficam bem em tabelas definidas. Outros são semiestruturados ou não estruturados e exigem outras formas de armazenamento e processamento. O tipo de dado influencia diretamente a tecnologia e o método usados para extrair valor.

Big Data não é só volume. É a combinação de escala, diversidade, velocidade, qualidade e propósito. Por isso, muda a forma como pensamos armazenamento, processamento e interpretação.

Estrutura dos Dados

Depois de entender valor, contexto e escala, precisamos olhar para a organização dos dados. Organizar dados não é só uma tarefa técnica; é uma forma de representar partes importantes de um negócio ou de uma realidade específica.

A forma como organizamos os dados influencia diretamente a capacidade de um sistema armazenar, consultar e relacionar informações. Uma estrutura inadequada pode dificultar análises, aumentar a complexidade do desenvolvimento e comprometer a qualidade dos dados.

Na modelagem de dados, essas organizações são chamadas de entidades. Uma entidade representa um conjunto de dados que descreve algo do mundo real dentro de um contexto. Cada entidade é composta por atributos, que representam suas características.

Se quisermos representar alunos, por exemplo, podemos definir atributos como nome, data de nascimento e CPF. Esses atributos, em conjunto, permitem identificar e descrever cada aluno dentro do sistema.

Identificar corretamente entidades e atributos é uma das primeiras etapas da modelagem de dados. Uma entidade mal definida pode gerar redundâncias, inconsistências ou dificuldades futuras na evolução do sistema.

A maneira como os dados são organizados, ou até a ausência dessa organização, determina sua classificação. Em geral, os dados podem ser classificados em três tipos: estruturados, semiestruturados e não estruturados.

Dados estruturados

Dados estruturados possuem um schema, ou esquema, fixo. Isso significa que todos os registros seguem o mesmo conjunto de campos previamente definidos. O schema funciona como um molde: determina quais atributos existem e qual formato devem seguir.

Essa rigidez pode parecer uma limitação, mas ajuda bastante. Como todos os registros seguem a mesma estrutura, fica mais simples

validar informações, criar relacionamentos e garantir a integridade dos dados.

Esse tipo de dado costuma ser representado em tabelas. Cada linha representa uma instância da entidade; cada coluna representa um atributo.

ID	NOME	CPF	DT NASCIMENTO
1	João Dantas	695.455.940-71	10/03/2000
2	Marcos Lima	326.475.060-00	10/03/2001
3	Mariana Dias	217.951.340-20	10/03/1990

Figura 1: Exemplo da entidade Aluno com atributos e instâncias

Observe que todos os registros seguem a mesma estrutura. Não é possível adicionar um aluno sem informar os campos definidos pelo *schema*.

Dados estruturados costumam ser armazenados em bancos de dados relacionais. Neles, as informações são organizadas em tabelas que podem se relacionar entre si por meio de chaves. Essas chaves identificam cada registro e permitem criar vínculos entre diferentes entidades, formando o chamado modelo relacional.

Esse modelo privilegia consistência, integridade e previsibilidade.

Por exemplo, em um sistema escolar, uma matrícula não deveria existir sem estar associada a um aluno. Bancos relacionais utilizam mecanismos que ajudam a garantir esse tipo de consistência.

Dados semiestruturados

Dados semiestruturados possuem um *schema* flexível. Existe organização, mas não há rigidez absoluta.

Diferentemente dos dados estruturados, nem todas as instâncias precisam ter exatamente os mesmos atributos. Um registro pode conter um campo adicional; outro pode omiti-lo.

Imagine uma entidade Aluno em que o atributo *email* não seja obrigatório. Um aluno pode ter email cadastrado; outro, não. Ambos continuam válidos.

Essa flexibilidade é característica de formatos como JSON (JavaScript Object Notation), muito usado na troca de dados entre

sistemas. O JSON foi popularizado por Douglas Crockford como uma forma simples e leve de representação.

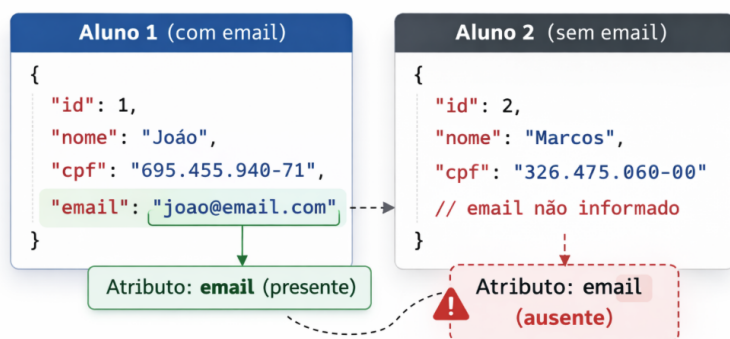


Figura 2: Exemplo de um JSON representando um Aluno

O segundo registro não possui o atributo email e, mesmo assim, continua válido. Se o dado estruturado lembra um formulário fixo, que todos precisam preencher da mesma forma,

o semiestruturado se parece mais com um formulário adaptável. Essa flexibilidade ajuda quando os registros têm características diferentes ou quando a estrutura muda com frequência.

Dados não estruturados

Existem também dados que não possuem *schema* explícito. São os chamados dados não estruturados.

Textos livres, áudios, vídeos, imagens e arquivos binários são exemplos. Eles não seguem uma organização formal que permita sua divisão direta em campos e colunas.

Um documento de texto pode parecer organizado para quem lê. Há parágrafos, títulos, seções. Para a máquina, porém, esse documento

é apenas uma sequência de caracteres. Não há, de forma explícita, campos definidos como nome, data ou valor.

Para extrair informação desse tipo de dado, são necessários processos adicionais, como indexação, processamento de linguagem natural, reconhecimento de padrões ou análise de mídia. Um currículo em PDF, por exemplo, pode conter nome, telefone e experiência profissional. Embora uma pessoa consiga identificar esses dados facilmente, um sistema precisará analisá-los para transformá-los em informações utilizáveis.



Figura 3: Exemplo de arquivos contendo dados não estruturados

Cada tipo de dado exige um cuidado diferente. Entender essa diferença ajuda a escolher como armazenar, consultar e manipular as informações. Depois disso, vem a pergunta prática: onde esses dados ficam guardados e como serão recuperados quando forem necessários?

O que é um Banco de Dados

Depois de entender o que são dados e informação, a pergunta seguinte é bem prática: onde tudo isso fica guardado com segurança?

A resposta está nos bancos de dados.

Para entender seu papel, imagine uma grande biblioteca pública. Se milhares de livros fossem simplesmente empilhados no chão de uma sala, encontrar uma informação específica seria uma tarefa demorada e frustrante.

Uma biblioteca funciona porque existe organização. Os livros são separados por categorias, identificados por etiquetas e

distribuídos em estantes de forma que possam ser encontrados rapidamente.

Um banco de dados segue o mesmo princípio no mundo digital. Ele é responsável por armazenar e organizar informações de maneira estruturada, permitindo que sejam localizadas, consultadas e atualizadas com facilidade.

Além de organizar os dados, as tecnologias de banco de dados ajudam a proteger as informações, manter sua integridade e garantir que elas estejam disponíveis quando necessário.

Graças a essa organização, sistemas conseguem localizar informações específicas em poucos instantes, mesmo quando lidam com milhares ou milhões de registros.

Sempre que você consulta o saldo de uma conta bancária, pesquisa um produto em uma loja virtual ou visualiza os dados de um cliente em um sistema, há um banco de dados trabalhando nos bastidores para fornecer essas informações.

Mas um banco de dados não existe apenas para armazenar informações. Seu papel também é organizar os dados de modo que possam se relacionar entre si.

Imagine um sistema de vendas. Nele existem clientes, produtos e vendas. Cada um desses elementos possui seus próprios dados, mas eles também se relacionam. Uma venda pertence a um cliente. Uma venda contém produtos. Um produto pode aparecer em diversas vendas. O banco de dados permite

representar essas relações de forma organizada, evitando duplicidade de informações e facilitando consultas futuras.

Sem um banco de dados, muitas aplicações dependeriam de arquivos isolados ou planilhas dispersas. À medida que a quantidade de dados aumentasse, surgiriam problemas como informações duplicadas, dificuldade de atualização e maior risco de inconsistências.

Por isso, bancos de dados se tornaram peças centrais dos sistemas modernos. Eles permitem armazenar grandes volumes de dados, manter relações entre informações e recuperar registros com rapidez e confiança.

Um banco de dados, então, é mais do que um lugar onde informações ficam guardadas. Ele organiza registros, preserva relações e

permite que dados sejam recuperados e usados por pessoas, empresas e sistemas.

Ciclo de Vida dos Dados

Até aqui falamos sobre estrutura e armazenamento. Mas existe uma etapa anterior: o momento em que o dado nasce. De onde ele vem? Em que ponto um acontecimento do mundo real passa a existir dentro de um sistema?

Antes de serem organizados em tabelas ou planilhas, os dados nascem de fatos do mundo real. Ações e eventos, quando registrados, se transformam em algo que podemos guardar e analisar. É nesse momento que um acontecimento passa a existir também dentro do sistema.

Para ilustrar, vamos usar um sistema de PDV (Ponto de Venda). Esse sistema é composto por vários módulos, como o cadastro de produtos e o registro de vendas. Cada ação realizada gera dados que registram o fato ocorrido.

Por exemplo, quando um produto é vendido, nascem dados como:

- Qual produto foi vendido.
- Quantidade vendida.
- Valor total da venda.
- Tipo de pagamento.
- Data e hora da venda.
- Funcionário que realizou a venda.
- Cliente que realizou a compra.

Esses dados só existem porque o evento foi registrado no sistema. Sem esse registro, a

venda teria ocorrido no mundo físico, mas não seria reconhecida pelo sistema. Com isso, qualquer análise ou decisão baseada nela se tornaria impossível.

Esse é o nascimento do dado: o momento em que um fato do mundo real vira registro digital.

Armazenamento

Depois que nascem, os dados precisam ser guardados e organizados para que possam ser usados. Aí entram o armazenamento e o tratamento.

No PDV, cada dado gerado pela venda de um produto ou pelo cadastro de clientes é armazenado em um banco de dados. O

armazenamento funciona como uma prateleira organizada, onde cada dado tem seu lugar.

Como vimos no capítulo anterior, essa organização é chamada de entidade: um conjunto estruturado de informações sobre um mesmo tipo de objeto, como produto, cliente ou venda. Exemplo:

- As informações de vendas ficam em tabelas de vendas.
- Os dados de produtos ficam em tabelas de produtos.
- Os dados de clientes ficam em tabelas de clientes.

Guardar dados é mais do que salvar em algum lugar. É garantir que eles possam ser recuperados com facilidade, sem se perder nem se misturar com informações de outro tipo.

Tratamento

Dados brutos nem sempre estão prontos para análise. Mesmo depois de armazenados, podem precisar de ajustes para ganhar qualidade e confiabilidade. Esse tratamento pode incluir:

- **Limpeza:** corrigir erros ou registros incompletos (como um preço incorreto ou uma data ausente).
- **Padronização:** organizar os dados em um formato uniforme (por exemplo, datas no mesmo padrão ou nomes escritos da mesma forma).
- **Validação:** verificar se os dados fazem sentido dentro do contexto do sistema (como impedir que a quantidade vendida seja negativa).

Parte da validação acontece antes mesmo de o dado ser salvo, no momento do cadastro. O sistema pode impedir que informações inválidas sejam registradas.

Ainda assim, mesmo com essas regras iniciais, pode ser necessário tratar os dados depois de armazenados, especialmente quando há registros antigos, importações de outros sistemas ou inconsistências que passaram despercebidas.

No PDV, antes de gerar relatórios ou indicadores, o sistema precisa garantir que cada registro de venda ou estoque esteja correto e padronizado. Um dado mal registrado pode distorcer a interpretação e levar a decisões ruins.

Análise de Dados

Quando os dados já foram registrados, armazenados e organizados, chega a hora de analisá-los. É aí que eles deixam de ser apenas registros e começam a responder perguntas. Em um PDV, por exemplo, relatórios podem mostrar:

- Vendas por vendedor no mês.
- Quem mais vendeu.
- Qual tipo de produto foi mais vendido (Top 10).
- Qual cliente mais comprou no mês.

Observe algo importante: essas informações não existem prontas no momento da venda. O sistema apenas registra:

- Produto;
- Quantidade;
- Valor;
- Vendedor;
- Cliente;
- Data.

O relatório é construído depois, a partir da combinação desses dados.

Esses relatórios não existem só para exibir números. Eles ajudam a responder perguntas do negócio.

Se um produto aparece constantemente entre os mais vendidos, a empresa pode aumentar o estoque. Se um vendedor apresenta desempenho acima da média, é possível identificar boas práticas e replicá-las na equipe. Se determinados clientes compram com

frequência, podem ser criadas ações de fidelização.

Os dados registrados durante as vendas continuam os mesmos. O que muda é a forma de organizá-los e interpretá-los. A análise transforma registros do passado em conhecimento útil para decisões do presente e para o planejamento do futuro.

Quem cuida dos Dados?

Depois de passar por nascimento, armazenamento, tratamento e análise, fica claro que os dados não percorrem esse caminho sozinhos.

Por trás desse processo existem profissionais com responsabilidades específicas. Cada um atua em uma parte do ciclo de vida dos dados.

À medida que empresas e sistemas passaram a produzir volumes cada vez maiores de dados, tornou-se necessário especializar funções. Alguns profissionais focam na captura

e no armazenamento; outros, na organização, análise e utilização estratégica.

Os nomes dos cargos podem mudar de empresa para empresa, mas algumas funções aparecem com frequência.

O Desenvolvedor de Software

Antes que qualquer dado possa ser armazenado ou analisado, é necessário existir um sistema capaz de registrá-lo.

O Desenvolvedor de Software é o profissional responsável por construir essas aplicações. É ele quem implementa formulários, regras de negócio, integrações e funcionalidades que transformam acontecimentos do mundo real em registros digitais.

Quando um cliente realiza uma compra em um sistema de PDV, por exemplo, é o software desenvolvido por esse profissional que captura informações como produto, quantidade, valor e data da venda.

Em muitos casos, o desenvolvedor também implementa validações que ajudam a garantir a qualidade dos dados desde o momento em que eles nascem.

Foco: Desenvolvimento de sistemas, regras de negócio e captura de dados.

No ciclo de vida: Atua principalmente no nascimento dos dados.

O DBA (Administrador de Banco de Dados)

Depois de capturados, os dados precisam ser armazenados de forma segura e organizada.

O DBA (Database Administrator) é o profissional responsável por administrar os bancos de dados utilizados pela organização. Sua função é garantir que as informações permaneçam disponíveis, seguras e acessíveis.

Ele também monitora o desempenho das consultas, realiza backups, controla permissões de acesso e adota medidas para proteger os dados contra falhas e perdas.

Se o banco de dados pode ser comparado a uma biblioteca, o DBA é quem mantém essa

biblioteca organizada, protegida e funcionando corretamente.

Foco: Administração, segurança, disponibilidade e desempenho dos bancos de dados.

No ciclo de vida: Atua principalmente no armazenamento e preservação dos dados.

O Engenheiro de Dados

À medida que os dados passam a ser produzidos por diversos sistemas, surge a necessidade de movimentá-los e integrá-los.

O Engenheiro de Dados é responsável por construir os mecanismos que coletam, transformam e distribuem dados entre diferentes ambientes.

Seu trabalho envolve a criação de processos automatizados que transportam informações de sistemas operacionais para ambientes analíticos, garantindo que os dados cheguem completos, consistentes e no momento adequado.

Simplificando, ele constrói as “estradas” por onde os dados trafegam dentro da organização.

Foco: Integração, movimentação, transformação e infraestrutura de dados.

No ciclo de vida: Atua principalmente na coleta, tratamento e movimentação dos dados.

O Analista de Dados

Depois que os dados foram coletados, organizados e preparados, eles podem ser analisados.

O Analista de Dados é o profissional responsável por transformar conjuntos de dados em informações compreensíveis para o negócio.

Ele produz relatórios, dashboards e indicadores capazes de responder perguntas importantes, como:

- Qual produto mais vendeu?
- Qual região gerou maior faturamento?
- Houve crescimento ou queda nas vendas?
- Quais clientes compram com maior frequência?

Seu trabalho permite que gestores e equipes compreendam o que aconteceu e tomem decisões mais fundamentadas.

Foco: Relatórios, indicadores, visualização de dados e inteligência de negócio.

No ciclo de vida: Atua principalmente na fase de análise.

O Cientista de Dados

Enquanto o analista busca compreender o que aconteceu, o Cientista de Dados procura identificar padrões e prever o que pode acontecer.

Para isso, utiliza estatística, matemática, aprendizado de máquina (Machine Learning) e técnicas avançadas de análise.

Seu objetivo é construir modelos capazes de encontrar relações complexas nos dados e gerar previsões ou recomendações.

Exemplos comuns incluem:

- Previsão de vendas futuras.
- Identificação de fraudes.
- Sistemas de recomendação.
- Previsão de inadimplência.
- Análise de comportamento de clientes.

O foco não está apenas em explicar o passado, mas em usar os dados para antecipar cenários e apoiar decisões futuras.

Foco: Modelagem estatística, previsão e Inteligência Artificial.

No ciclo de vida: Atua na extração de conhecimento e geração de previsões.

Trabalhando em Conjunto

Essas funções têm responsabilidades diferentes, mas fazem parte do mesmo processo.

Os dados podem nascer em um sistema desenvolvido por um programador, ser armazenados em bancos administrados por DBAs, movimentados por engenheiros de dados, analisados por analistas e utilizados por cientistas de dados na construção de modelos preditivos.

Em organizações menores, uma única pessoa pode acumular várias dessas responsabilidades. Em empresas maiores, equipes especializadas trabalham de forma integrada para garantir que os dados percorram

todo o ciclo de vida com qualidade, segurança e utilidade.

No fim, independentemente do cargo ou da tecnologia, todos trabalham para que dados se transformem em informação, conhecimento e decisões melhores.

Qualidade e Ética

O ciclo de vida dos dados parece simples quando olhamos de longe: o dado nasce, é armazenado, tratado, analisado e usado em decisões. Na prática, esse caminho só funciona bem quando a qualidade é preservada em cada etapa.

Um dado incorreto na origem compromete todo o restante do processo. Se uma venda é registrada com valor errado, o relatório mensal refletirá esse erro. Se a quantidade de um produto for digitada incorretamente, o controle de estoque será afetado. O problema não estará no relatório final, mas no momento em que o dado nasceu.

Erros não são o único problema.

Dados incompletos limitam a análise. Se o tipo de pagamento não for registrado, não será possível saber qual forma de pagamento é mais usada. Se o cliente não for identificado, não será possível descobrir quem mais compra durante o mês. Perguntas deixam de ser respondidas não por falta de análise, mas por falta de registro adequado.

Dados inconsistentes também geram dificuldades. Quando informações semelhantes são registradas de maneiras diferentes, como categorias de produtos com nomes variados para representar o mesmo item, o tratamento se torna mais complexo e a análise pode apresentar distorções.

Por isso, qualidade não significa apenas “estar certo”. Também significa ser completo, consistente e confiável.

A responsabilidade pelo uso adequado dos dados é compartilhada. Quem registra precisa fazê-lo com atenção. Quem organiza e trata deve validar e padronizar. Quem analisa precisa questionar possíveis falhas. E quem decide deve compreender que toda decisão baseada em dados depende da qualidade do que foi coletado. Essa responsabilidade também é ética: dados mal coletados, mal tratados ou usados sem critério podem prejudicar pessoas e decisões.

Os dados transmitem uma aparência de precisão. Números parecem objetivos. Relatórios parecem definitivos. Mas, se a base

estiver comprometida, a conclusão também estará.

Compreender o ciclo de vida dos dados também é reconhecer a responsabilidade envolvida em cada etapa. Quando são bem registrados, organizados e analisados, os dados apoiam decisões melhores. Quando são negligenciados, podem produzir interpretações equivocadas e escolhas injustas ou mal fundamentadas.

O valor dos dados depende da forma como eles são produzidos e utilizados.

LGPD

Até aqui falamos sobre dados como registros, estruturas e base para decisões. Falta olhar para um ponto essencial: muitos dados dizem respeito a pessoas.

Quando isso acontece, a discussão deixa de ser apenas técnica. Entra em cena o direito das pessoas sobre as informações que dizem respeito a elas.

Quando uma empresa registra o nome de um cliente, seu CPF, endereço, telefone ou e-mail, ela não passa a ser dona dessas informações. Ela assume a responsabilidade de armazená-las e utilizá-las de forma adequada.

Com sistemas digitais coletando cada vez mais dados pessoais, tornou-se necessário criar regras para proteger a privacidade dos cidadãos e limitar usos indevidos.

Foi nesse contexto que surgiu a Lei Geral de Proteção de Dados Pessoais, mais conhecida como LGPD.

O que é a LGPD?

A Lei Geral de Proteção de Dados Pessoais (LGPD), instituída pela Lei nº 13.709/2018, estabelece regras sobre como dados pessoais podem ser coletados, armazenados, utilizados, compartilhados e eliminados.

Seu principal objetivo é garantir que os titulares tenham mais controle sobre suas

informações pessoais e que organizações públicas e privadas atuem com transparência e responsabilidade ao tratar esses dados.

A LGPD busca responder perguntas como:

- Quais dados podem ser coletados?
- Por que esses dados estão sendo coletados?
- Quem terá acesso a eles?
- Por quanto tempo serão armazenados?
- Como serão protegidos?
- O titular pode solicitar sua exclusão?

A lei tenta equilibrar duas necessidades: permitir que organizações usem dados de forma legítima e proteger os direitos dos cidadãos.

O que são dados pessoais?

A LGPD se aplica principalmente aos chamados dados pessoais.

Dados pessoais são informações que identificam ou podem identificar uma pessoa, direta ou indiretamente.

Alguns exemplos são:

- Nome.
- CPF.
- RG.
- E-mail.
- Telefone.
- Endereço.
- Localização geográfica.
- Dados de acesso a sistemas.

Dependendo do contexto, até mesmo um conjunto de informações aparentemente simples pode permitir a identificação de uma pessoa.

A lei também define uma categoria especial chamada dados pessoais sensíveis, que exigem proteção adicional. Entre eles estão informações relacionadas à saúde, biometria, origem racial ou étnica, convicções religiosas e outros dados capazes de gerar discriminação ou impactos significativos na vida do indivíduo.

Breve Histórico

A LGPD foi sancionada em 14 de agosto de 2018 e entrou em vigor em setembro de 2020, com aplicação de sanções administrativas a partir de agosto de 2021.

A lei foi fortemente inspirada no GDPR (General Data Protection Regulation), regulamento europeu de proteção de dados que se tornou referência internacional em privacidade e tratamento de informações pessoais.

Sua criação acompanhou uma preocupação cada vez maior com o uso de dados por empresas, governos e plataformas digitais.

Os Princípios da LGPD

A LGPD estabelece princípios para orientar o tratamento de dados pessoais. Eles servem como referência para qualquer

organização que utilize informações de cidadãos.

Finalidade e Adequação

Os dados devem ser utilizados para finalidades legítimas, específicas e previamente informadas ao titular.

O uso realizado também deve ser compatível com a finalidade originalmente apresentada.

Se uma empresa solicita um e-mail para envio de uma nota fiscal, não deve usá-lo depois em campanhas de marketing sem uma justificativa adequada ou outra base legal aplicável.

Necessidade

A coleta deve se limitar ao mínimo necessário para atingir a finalidade proposta.

Isso significa que organizações não devem solicitar informações que não tenham relação direta com o serviço prestado.

Quanto menos dados forem coletados, menor será o risco associado ao tratamento deles.

Livre Acesso e Qualidade dos Dados

Os titulares têm o direito de saber quais informações estão armazenadas sobre eles.

Os dados também devem ser corretos, atualizados e completos.

Informações incorretas podem gerar prejuízos tanto para o titular quanto para a organização que as utiliza.

Transparência

As pessoas devem compreender com clareza como seus dados estão sendo utilizados.

Políticas de privacidade e termos de uso devem ser acessíveis e compreensíveis, evitando linguagem excessivamente técnica ou ambígua.

Segurança e Prevenção

As organizações devem adotar medidas técnicas e administrativas para proteger os dados contra acessos não autorizados, vazamentos, perdas ou alterações indevidas.

A segurança não deve aparecer apenas como reação a incidentes, mas como uma atividade preventiva e contínua.

Não Discriminação

Os dados não podem ser utilizados para práticas discriminatórias, abusivas ou ilegais.

Decisões automatizadas baseadas em informações pessoais devem respeitar os direitos fundamentais dos indivíduos.

Responsabilização e Prestação de Contas

As organizações devem ser capazes de demonstrar que cumprem as exigências da legislação.

Não basta afirmar que os dados estão protegidos; é necessário ter evidências,

processos e controles que comprovem essa proteção.

LGPD e o Ciclo de Vida dos Dados

Os conceitos do ciclo de vida ajudam a enxergar a LGPD de forma prática.

Quando os dados nascem, é necessário justificar sua coleta.

Quando são armazenados, é necessário protegê-los.

Quando são tratados, é necessário garantir sua qualidade e consistência.

Quando são compartilhados, é necessário controlar quem terá acesso às informações.

Quando são analisados, é necessário respeitar os direitos dos titulares e evitar usos inadequados.

A LGPD não aparece só no começo ou no fim. Ela acompanha todo o ciclo de vida dos dados.

Proteger informações deixa de ser uma tarefa isolada e passa a ser uma responsabilidade permanente.

O Papel da ANPD

Para fiscalizar o cumprimento da LGPD, foi criada a Autoridade Nacional de Proteção de Dados (ANPD).

A ANPD, atualmente Agência Nacional de Proteção de Dados, é responsável por orientar

organizações, fiscalizar o cumprimento da legislação e aplicar sanções quando necessário.

Entre suas atribuições estão a elaboração de diretrizes, a análise de incidentes de segurança e a promoção de boas práticas relacionadas à proteção de dados pessoais.

A LGPD lembra que dados não são apenas ativos tecnológicos ou recursos de negócio. Em muitos casos, representam aspectos da identidade das pessoas.

Por isso, organizações não podem pensar apenas em coletar, armazenar e analisar informações. Também precisam respeitar os direitos das pessoas às quais esses dados pertencem.

Em um mundo orientado por dados, privacidade e proteção da informação precisam caminhar junto com bancos de dados, algoritmos e ferramentas analíticas.

A tecnologia aumenta nossa capacidade de coletar e processar informações. A LGPD lembra que essa capacidade precisa vir acompanhada de responsabilidade.

Considerações Finais

Dados não são apenas números guardados em sistemas. Eles nascem de fatos, são registrados, organizados, tratados e analisados até se transformarem em informação capaz de orientar decisões.

Dados isolados não explicam a realidade. Eles precisam de contexto, estrutura e interpretação. Quantidade não substitui qualidade. Grandes volumes podem ampliar a capacidade de análise, mas também podem aumentar ruídos, erros e interpretações equivocadas.

Os dados podem assumir diferentes estruturas, percorrer um ciclo de vida e exigir

cuidado em cada etapa. Um erro na origem compromete o resultado final. Uma falha na organização limita a análise. Uma interpretação descuidada pode gerar decisões inadequadas.

No ambiente digital, decisões estratégicas, operacionais e até sociais dependem de registros armazenados em sistemas. Isso aumenta a importância da modelagem, da qualidade e da responsabilidade no uso das informações.

Dados não substituem julgamento humano, mas influenciam escolhas. Eles não eliminam incertezas; quando bem utilizados, ajudam a reduzi-las.

Compreender dados é compreender como parte da realidade entra nos sistemas. Essa tradução nunca é neutra. Depende de escolhas,

critérios, qualidade e responsabilidade. Aprender a olhar para os dados com cuidado é um passo importante para construir sistemas melhores e tomar decisões mais conscientes.

Referências

AMARAL, Fernando. *Aprenda Mineração de Dados*.

ALVES, William Pereira. *Análise e Projeto de Sistemas*.

GONICK, Larry. *Introdução Ilustrada à Computação*.