

UNISENAI  
MURILO CÉSAR FERREIRA

**BigData e Analytics:**  
Fundamentos e Aplicações

JOINVILLE

2025

## **Introdução**

O avanço tecnológico e a crescente digitalização dos processos em diversos setores têm impulsionado uma produção exponencial de dados, criando um cenário no qual o volume e a complexidade das informações ultrapassam a capacidade das ferramentas tradicionais de gerenciamento. Esse contexto é caracterizado pelo fenômeno conhecido como Big Data, que envolve o manejo de conjuntos de dados em escala massiva, produzidos em velocidades elevadas e com elevada diversidade de formatos. O estudo e a compreensão desse conceito tornam-se essenciais para o desenvolvimento de estratégias eficazes de armazenamento, processamento e análise, capazes de extrair valor dessas informações e atender às demandas contemporâneas de diferentes áreas.

A relevância do Big Data está associada não apenas à quantidade de dados gerados, mas também à diversidade de suas origens e formatos, o que impõe desafios significativos para a infraestrutura tecnológica e os métodos analíticos empregados. Ao mesmo tempo, a qualidade e a confiabilidade das informações coletadas influenciam diretamente a assertividade das análises e das decisões tomadas com base nesses dados. Assim, entender os atributos que compõem o Big Data — incluindo volume, velocidade, variedade, veracidade e valor — é fundamental para estabelecer bases sólidas na gestão da informação e na transformação dos dados em conhecimento útil.

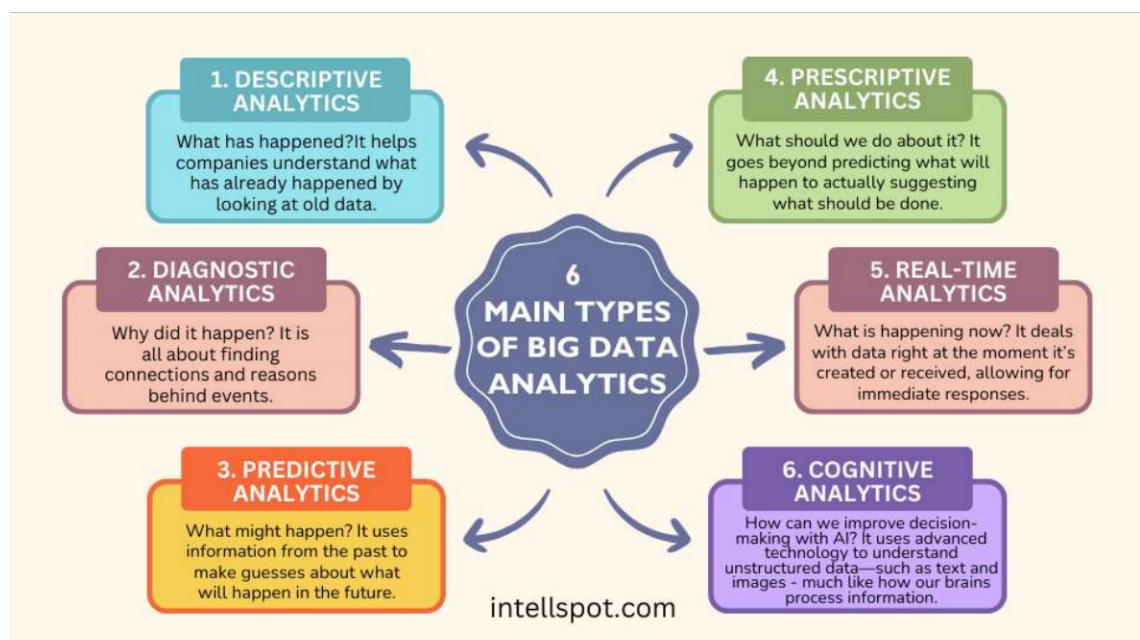
No âmbito da análise de dados, destaca-se a distinção conceitual entre Big Data e Analytics, embora essas áreas sejam interligadas e complementares. Enquanto o Big Data diz respeito ao conjunto de dados em si e aos desafios inerentes à sua manipulação, a Analytics refere-se às técnicas e processos utilizados para interpretar esses dados, identificar padrões, antecipar comportamentos e apoiar a tomada de decisão. Essa diferenciação é crucial para compreender o papel estratégico que cada elemento desempenha no ciclo de transformação da informação, desde a coleta até a aplicação prática dos resultados obtidos.

Além disso, a aplicação prática do Big Data e da Analytics tem sido expandida em múltiplos setores, promovendo inovações e melhorias que impactam diretamente a eficiência, a produtividade e a qualidade dos serviços prestados. Setores como saúde, educação, indústria, transporte e comércio vêm incorporando essas tecnologias para otimizar processos, personalizar atendimentos, prever tendências e apoiar estratégias gerenciais. A convergência entre ferramentas especializadas, métodos analíticos

avançados e ambientes de dados robustos tem possibilitado uma nova dinâmica na forma como as organizações lidam com informações, transformando dados brutos em decisões fundamentadas e valor agregado.

## Big Data

Big Data é um termo utilizado para designar grandes conjuntos de dados, caracterizados por um volume, velocidade e variedade tão elevados que as ferramentas tradicionais de armazenamento, processamento e análise se tornam insuficientes. Esse conceito é frequentemente explicado por meio dos chamados “5 Vs”: volume, velocidade, variedade, veracidade e valor. O volume refere-se à quantidade massiva de dados gerados diariamente por diferentes fontes, como redes sociais, sensores e transações financeiras. A velocidade diz respeito à rapidez com que esses dados são produzidos e precisam ser processados, muitas vezes em tempo real. A variedade relaciona-se à diversidade de formatos, englobando dados estruturados, semiestruturados e não estruturados. A veracidade indica o grau de confiabilidade e qualidade dessas informações, sendo essencial para análises assertivas. Por fim, o valor está associado à capacidade de extrair informações relevantes e aplicáveis, que tragam benefícios concretos para organizações e indivíduos.

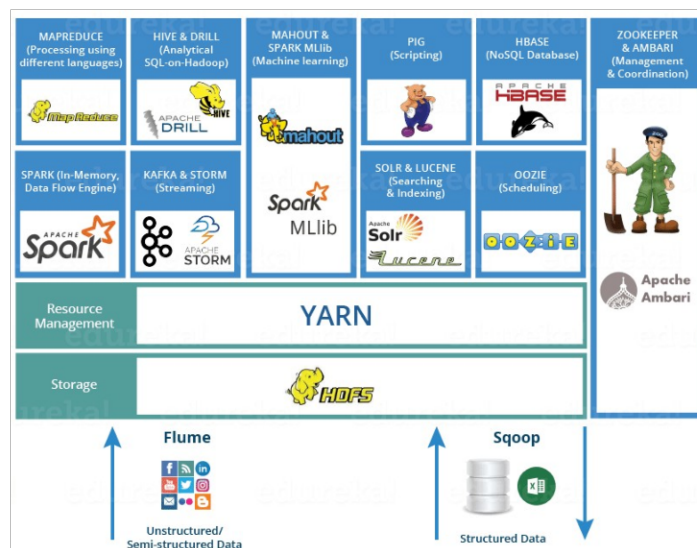


Embora frequentemente confundidos, Big Data e Analytics não são sinônimos. Big Data refere-se aos dados em si, ao desafio de lidar com grandes volumes e diferentes tipos

de informações, enquanto Analytics (ou análise de dados) diz respeito ao conjunto de métodos e técnicas utilizados para extrair padrões, prever comportamentos e apoiar a tomada de decisão a partir desses dados. Em outras palavras, Big Data é a matéria-prima, e Analytics é o processo que transforma essa matéria-prima em conhecimento útil.

No campo da análise de dados, identificam-se quatro tipos principais de abordagens. A análise descritiva busca responder à pergunta “o que aconteceu?”, utilizando-se de dados históricos para gerar relatórios, gráficos e painéis que resumem informações passadas. A análise diagnóstica procura entender “por que aconteceu?”, explorando correlações, identificando causas e avaliando padrões por meio de técnicas mais aprofundadas. A análise preditiva concentra-se em prever “o que provavelmente acontecerá?”, utilizando modelos estatísticos e algoritmos de aprendizado de máquina para antecipar tendências. Por fim, a análise prescritiva busca indicar “o que deve ser feito?”, oferecendo recomendações e simulando cenários para apoiar decisões estratégicas.

Para viabilizar o uso do Big Data, existe um conjunto diversificado de ferramentas especializadas. Entre as mais relevantes, destaca-se o Apache Hadoop, que oferece um sistema de armazenamento e processamento distribuído, e o Apache Spark, que realiza análises em lote e em tempo real com processamento em memória. Ferramentas como Apache Kafka e Apache Storm são amplamente empregadas para o tratamento de fluxos de dados em tempo real, enquanto o Apache Hive e o Apache Impala permitem consultas utilizando linguagem SQL sobre grandes volumes de dados armazenados em sistemas distribuídos. No campo da visualização e integração, soluções como Tableau, Power BI e Pentaho possibilitam criar dashboards interativos e análises visuais acessíveis,



favorecendo a interpretação dos resultados por usuários de diferentes áreas. Outras ferramentas, como Splunk, KNIME e RapidMiner, são aplicadas em análise de logs, modelagem preditiva e mineração de dados, ampliando as possibilidades de exploração e aplicação das informações coletadas.

As aplicações do Big Data se estendem a diferentes setores da sociedade. Na saúde, ele é utilizado para prever epidemias, personalizar tratamentos, analisar dados genômicos e otimizar a gestão hospitalar. Na educação, possibilita o monitoramento do desempenho acadêmico e a personalização de estratégias de ensino, a partir da análise do comportamento e do progresso dos estudantes. Na indústria, auxilia na manutenção preditiva de máquinas, na otimização de processos produtivos e na gestão eficiente de cadeias de suprimento. No setor de transporte e logística, permite a previsão de tráfego, o planejamento de rotas e o controle em tempo real de frotas. No comércio e marketing, é usado para segmentar clientes, criar recomendações personalizadas e conduzir campanhas mais assertivas. Esses exemplos demonstram como a combinação entre Big Data e Analytics pode gerar valor prático e estratégico, transformando dados brutos em decisões fundamentadas.

## **Aplicação de Big Data e Analytics no Setor de Saúde — Cenário de Prevenção de Doenças Crônicas**

### **Cenário escolhido**

O setor escolhido para aplicação dos conceitos de Big Data e Analytics é a saúde pública, com foco em um programa de prevenção e monitoramento de doenças crônicas não transmissíveis (DCNT), como diabetes tipo 2 e hipertensão. Trata-se de um contexto realista em que governos, hospitais e clínicas buscam utilizar dados em larga escala para melhorar a prevenção, reduzir custos e aumentar a qualidade de vida da população.

### **Problema ou oportunidade identificada**

Um dos principais desafios enfrentados por sistemas de saúde é o aumento contínuo de casos de DCNT, que representam um custo elevado para os cofres públicos e para operadoras de saúde privada. A ausência de monitoramento preventivo, aliada a diagnósticos tardios, eleva a necessidade de internações, exames de alto custo e tratamentos prolongados. A aplicação de Big Data e Analytics oferece a oportunidade de

criar modelos preditivos capazes de identificar indivíduos com maior risco de desenvolver essas doenças, permitindo intervenções precoces e personalizadas.

### **Fontes de dados possíveis**

Para viabilizar essa solução, diversas fontes de dados podem ser exploradas, incluindo:

- Registros eletrônicos de saúde (RES) contendo histórico médico, resultados de exames laboratoriais e prescrições.
- Dados de dispositivos vestíveis (wearables), como smartwatches e pulseiras de monitoramento, registrando frequência cardíaca, nível de atividade física e padrões de sono.
- Bases de dados públicas de saúde e estatísticas demográficas, como as fornecidas pelo IBGE e pelo DataSUS.
- Informações socioeconômicas e ambientais que possam influenciar hábitos de vida, como alimentação e acesso a serviços de saúde.

A obtenção desses dados pode ocorrer via integração com APIs de dispositivos, parcerias com laboratórios e clínicas, bem como coleta direta por meio de questionários eletrônicos.

### **Tipo de análise mais adequada**

A análise preditiva é a mais indicada para esse cenário, pois permite estimar a probabilidade de um indivíduo desenvolver uma DCNT com base em seu histórico e estilo de vida. Com isso, o sistema poderia identificar padrões e correlacionar variáveis, como índice de massa corporal (IMC), níveis de glicose e frequência de exercícios, determinando o grau de risco de cada paciente. Além disso, elementos de análise prescritiva poderiam ser incorporados para recomendar planos de ação específicos, como ajustes na dieta, aumento da atividade física ou realização de exames preventivos.

### **Ferramentas recomendadas**

Para a implementação do projeto, recomenda-se um conjunto integrado de ferramentas:

- Hadoop e Apache Spark para processamento e análise em larga escala.
- Python com bibliotecas como Pandas, Scikit-learn e TensorFlow para modelagem preditiva e machine learning.

- Power BI ou Tableau para visualização interativa dos dados e dashboards de acompanhamento.
- PostgreSQL para armazenamento e gerenciamento de dados estruturados.
- APIs REST para integração entre dispositivos vestíveis, sistemas de clínicas e a base central de dados.

### **Resultados esperados**

A adoção desse sistema de Big Data e Analytics traria benefícios expressivos, como:

- Redução de custos com internações e tratamentos de doenças crônicas, por meio de ações preventivas e diagnósticos precoces.
- Aumento da eficiência no uso de recursos médicos, priorizando atendimentos para pacientes de maior risco.
- Melhoria na qualidade de vida da população, ao permitir intervenções personalizadas e baseadas em evidências.
- Produção de relatórios estratégicos para gestores públicos e privados, apoiando decisões sobre políticas de saúde.
- Capacidade de aprendizado contínuo do sistema, aprimorando modelos preditivos a partir de novos dados coletados.

Em síntese, a aplicação de Big Data e Analytics no monitoramento de doenças crônicas no setor de saúde demonstra como o uso inteligente de dados pode gerar impactos positivos não apenas na economia, mas também no bem-estar coletivo, alinhando-se às diretrizes de saúde preventiva e medicina baseada em evidências.

### **Referências**

CETAX. *Data Science vs Big Data vs Data Analytics*. Disponível em: <https://cetax.com.br/data-science-vs-big-data-vs-data-analytics/>. Acesso em: 9 ago. 2025.

CIDESP. *Big Data: significado*. Disponível em: <https://cidesp.com.br/artigo/big-data-significado/>. Acesso em: 9 ago. 2025.

FAROL BI. *Conheça os 4 tipos de análise do Big Data Analytics*. Disponível em: <https://farolbi.com.br/conheca-os-4-tipos-de-analise-do-big-data-analytics/>. Acesso em: 9 ago. 2025.

IBM. *Big Data*. Disponível em: <https://www.ibm.com/br-pt/think/topics/big-data>. Acesso em: 9 ago. 2025.

MINERS, Mind. *Ferramentas de Big Data*. Disponível em: <https://mindminers.com/blog/ferramentas-de-big-data/>. Acesso em: 9 ago. 2025.

NAÇÃO DIGITAL. *Ferramentas de Big Data*. Disponível em: <https://nacao.digital/blog/ferramentas-de-big-data/>. Acesso em: 9 ago. 2025.

SALESFORCE. *Big Data*. Disponível em: <https://www.salesforce.com/br/blog/big-data/>. Acesso em: 9 ago. 2025.

INTELLSPOT. *Big Data vs Data Analytics: What's the Difference?*. Disponível em: <https://www.intellspot.com/big-data-vs-data-analytics/>. Acesso em: 9 ago. 2025.

EDUREKA. *Hadoop Ecosystem: Hadoop Components and Architecture Explained*. Disponível em: <https://www.edureka.co/blog/hadoop-ecosystem/>. Acesso em: 9 ago. 2025.