





# **FUNDAMENTOS DE BIG DATA**

O que é exatamente [big data](https://www.oracle.com/br/big-data/)?

Para realmente compreender o big data, é recomendável conhecer um pouco de sua história. Aqui está a definição do Gartner, criada por volta de 2001 (que ainda é considerada a definição mais confiável): Big data são dados com maior variedade que chegam em volumes crescentes e com velocidade cada vez maior. Isso é conhecido como os três Vs.

Simplificando, big data é um conjunto de dados maior e mais complexo, especialmente de novas fontes de dados. Esses conjuntos de dados são tão volumosos que o software tradicional de processamento de dados simplesmente não consegue gerenciá-los. No entanto, esses grandes volumes de dados podem ser usados para resolver problemas de negócios que você não conseguiria resolver antes.

## Os três Vs do big data

|  |  |
| --- | --- |
| **Volume** | A quantidade de dados importa. Com o big data, você terá que processar grandes volumes de dados não estruturados de baixa densidade. Podem ser dados de valor desconhecido, como feeds de dados do Twitter, fluxos de cliques em uma página da web ou em um aplicativo para dispositivos móveis, ou ainda um equipamento habilitado para sensores. Para algumas empresas, isso pode utilizar dezenas de terabytes de dados. Para outras, podem ser centenas de petabytes. |
| **Velocidade** | Velocidade é a taxa mais rápida na qual os dados são recebidos e talvez administrados. Normalmente, a velocidade mais alta dos dados é transmitida diretamente para a memória, em vez de ser gravada no disco. Alguns produtos inteligentes habilitados para internet operam em tempo real ou quase em tempo real e exigem avaliação e ação em tempo real. |
| **Variedade** | Variedade refere-se aos vários tipos de dados disponíveis. Tipos de dados tradicionais foram estruturados e se adequam perfeitamente a um banco de dados relacional. Com o aumento de big data, os dados vêm em novos tipos de dados não estruturados. Tipos de dados não estruturados e semiestruturados, como texto, áudio e vídeo, exigem um pré-processamento adicional para obter significado e dar suporte a metadados. |

# **FERRAMENTAS DE BIG DATA**

Agora que já entendeu o conceito e como funciona, deve estar se perguntando sobre qual ferramenta escolher. Acertei?

Veja bem, existem centenas de soluções disponíveis hoje no mercado.

Mas para não frustrar suas expectativas, preparei uma lista com o top 10 principais ferramentas de Big Data para os negócios.

Vamos a ela:

[Apache Hadoop](https://hadoop.apache.org/) – para aumentar ou diminuir o tamanho de qualquer arquivo

[Chartio](https://chartio.com/) – para fazer a combinação de dados coletados e criar relatórios

[Import.io](https://www.import.io/) – para extrair dados open source

[MindMiners](https://mindminers.com/platform) – para fazer pesquisas automatizadas

[Oracle Data Mining](https://www.oracle.com/marketingcloud/products/) – para peneirar os dados mais relevantes

[Pentaho](https://www.hitachivantara.com/go/pentaho.html?source=pentaho-redirect) – para integrar informações de diferentes fontes

[Statwing](http://www.statwing.com/) – para análise estatística

[STRATWs ONE](https://www.siteware.com.br/nossas-solucoes/) – para fazer a gestão de performance corporativa

[Tableau](https://www.tableau.com/) – para visualizar dados, criar gráficos, mapas e tabelas diversas

[Watson Analytics](https://www.ibm.com/watson-analytics) – para encontrar insights sobre os dados.

Para escolher a melhor ferramenta para o seu negócio, procure se atentar à sua interface, além de conhecer as funcionalidades, as atualizações e as restrições de segurança.

Assim, fica mais fácil escolher aquela que melhor atenda às suas necessidades corporativas.

Se preferir, salve este artigo nos seus favoritos e vá testando cada uma delas para ver com quais melhor se adapta.