



*Fundamentos de Engenharia de Dados*

# Fundamentos de Engenharia de Dados

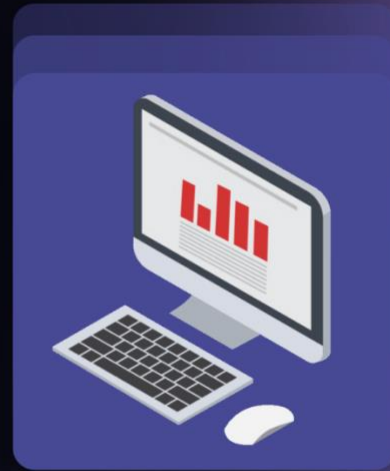
## Compressão de Arquivos Para o Armazenamento Distribuído



# Fundamentos de Engenharia de Dados

Curso Online, Gratuito, Em Português e com Certificado de Conclusão

[www.datascienceacademy.com.br](http://www.datascienceacademy.com.br)



A compressão é uma técnica comum usada em sistemas de armazenamento distribuído para reduzir a quantidade de espaço de armazenamento necessária para um determinado conjunto de dados. Isso pode ser especialmente útil em sistemas distribuídos de grande escala, onde o volume de dados armazenados pode ser muito grande.

Existem várias maneiras de usar a compactação em sistemas de armazenamento distribuído:

**No nível do arquivo:** Arquivos individuais podem ser compactados antes de serem armazenados no sistema de armazenamento distribuído. Isso pode ser feito usando vários algoritmos de compactação, como gzip ou bzip2, que podem reduzir significativamente o tamanho do arquivo.

**No nível do bloco:** Em alguns sistemas de armazenamento distribuído, os dados são armazenados em blocos, que normalmente são de tamanho fixo (por exemplo, 128 MB). Esses blocos podem ser compactados antes de serem armazenados, usando um algoritmo de compactação sem perdas ou com perdas.

**Compactação em trânsito:** Os dados também podem ser compactados à medida que são transferidos entre computadores no sistema distribuído. Isso pode ser feito usando uma variedade de técnicas, como compactação HTTP ou dimensionamento de janela TCP.



É importante observar que a eficácia da compactação dependerá das características dos dados que estão sendo compactados. Alguns tipos de dados, como texto ou código, podem ser altamente compactáveis, enquanto outros tipos de dados, como imagens ou vídeos, podem ser menos compactáveis.

Sendo assim, os dados podem ser armazenados em bancos de dados SQL, em bancos de dados NoSQL, em Data Lakes com diferentes formatos de arquivos e ainda é possível aplicar compressão. A decisão da arquitetura de armazenamento depende do volume de dados, do padrão de acesso, tipo de processamento e casos de uso.

Para conhecer nossos demais cursos e formações, acesse o link abaixo:

<https://www.datascienceacademy.com.br/todoscursosdsa>