



*Fundamentos de Engenharia de Dados*

# Fundamentos de Engenharia de Dados

## Como IaC é Usada em Engenharia de Dados?

# 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 Infraestrutura como código (IaC) é uma prática comum na Engenharia de Dados, pois pode ser usada para gerenciar e automatizar a criação e configuração de toda a infraestrutura necessária para processar, armazenar e analisar grandes volumes de dados.

Algumas das maneiras como a IaC é usada na Engenharia de Dados incluem:

**Criação de ambientes de infraestrutura:** A IaC pode ser usada para criar ambientes de infraestrutura de dados, incluindo clusters de processamento distribuído, bancos de dados, armazenamento em nuvem, entre outros. Esses ambientes podem ser criados rapidamente e de forma consistente, garantindo que todos os ambientes estejam configurados de maneira idêntica e evitando problemas de configuração manual.

**Configuração de serviços de processamento de dados:** A IaC pode ser usada para configurar serviços de processamento de dados, como o Apache Spark ou Apache Flink. As configurações desses serviços podem ser mantidas em código e versionadas, permitindo a colaboração e a rastreabilidade.

**Gerenciamento de fluxos de dados:** A IaC pode ser usada para configurar e gerenciar fluxos de dados que movem dados de um ambiente para outro. Os códigos de infraestrutura podem ser usados para automatizar a criação e configuração de pipelines de dados, bem como para atualizar e modificar esses pipelines à medida que as necessidades de dados mudam.

**Configuração de monitoramento e alertas:** A IaC pode ser usada para configurar o monitoramento de métricas de dados e infraestrutura, como uso de CPU, uso de disco, latência

---

de rede, entre outros. Os alertas podem ser configurados para notificar automaticamente as equipes de operações quando as métricas se desviam dos valores desejados.

A IaC está se tornando uma prática essencial na Engenharia de Dados, pois permite que a infraestrutura de dados seja gerenciada como código, reduzindo o tempo e o esforço necessários para configurar e gerenciar infraestrutura complexa e permitindo que as equipes se concentrem na análise e processamento de dados em si.