

Google Dataflow: Construção e Operação de Pipelines de Dados

Curso: Google Dataflow - Construção e Operação de Pipelines de Dados
Carga Horária: 26 horas (16 horas ao vivo + 10 horas atividades orientadas)
Categorias: Engenharia de Dados
Nível: Intermediário
Plataforma: Google Cloud
Especialista: A definir
Formato: Ao vivo

Sobre o curso

À medida que pipelines de dados passam a sustentar sistemas analíticos e operacionais, surge a necessidade de construí-los, operá-los e mantê-los em funcionamento contínuo, lidando com volumes maiores, dados em tempo real e falhas que impactam diretamente o negócio. Nesse ponto, não basta entender conceitos — é preciso executar, monitorar e tomar decisões técnicas.

O curso Google Dataflow – Construção e Operação de Pipelines de Dados foi desenvolvido para profissionais que já conhecem os fundamentos do Dataflow e precisam atuar de forma mais autônoma na criação e operação de pipelines batch e streaming, integrando diferentes fontes e destinos de dados no Google Cloud.

Com foco prático, o curso aborda a construção de pipelines mais robustos, o uso de conceitos como janelas, estado e integração com serviços como BigQuery, Pub/Sub e Cloud Storage, além do monitoramento e operação de pipelines em ambientes produtivos.

O participante desenvolve a capacidade de identificar problemas, interpretar métricas operacionais e ajustar o funcionamento dos pipelines, habilidades essenciais para atuação como Data Engineer Pleno.

Módulos do Curso

Módulo 1 — Arquitetura de Pipelines no Dataflow

- Desenho de pipelines escaláveis
- Integração com outros serviços GCP

Módulo 2 — Apache Beam na Prática

- Transforms avançadas
- Combinações, agregações e joins

Módulo 3 — Processamento de Streaming

- Janelas, watermarks e triggers
- Estado em pipelines
- Processamento contínuo em tempo real

Módulo 4 — Integração com Fontes e Destinos

- BigQuery, Pub/Sub, Cloud Storage
- Escrita eficiente de dados

Módulo 5 — Operação e Monitoramento

- Monitoramento de pipelines
- Tratamento de falhas
- Ajustes iniciais de desempenho

O que você vai aprender?

- Construção de pipelines batch e streaming no Dataflow
- Uso avançado do Apache Beam
- Processamento de dados em tempo real
- Integração com serviços de dados do Google Cloud
- Monitoramento e operação de pipelines produtivos
- Atuação prática como Data Engineer Pleno

Pra quem é o curso?

Este curso é indicado para:

- Data Engineers em nível Pleno
- Profissionais atuando com pipelines de dados
- Desenvolvedores que trabalham com dados em escala
- Pessoas em preparação para certificação Google Data Engineer

Pré-requisitos

- Conhecimento prévio de Google Dataflow ou curso Iniciante
- Programação em Python ou Java

— Noções de pipelines de dados e arquitetura analítica

rox