

Google Dataflow: Fundamentos de Pipelines de Dados

Curso: Google Dataflow - Fundamentos de Pipelines de Dados
Carga Horária: 24 horas (16 horas ao vivo + 8 horas atividades orientadas)
Categorias: Engenharia de Dados
Nível: Iniciante
Plataforma: Google Cloud
Especialista: A definir
Formato: Ao vivo

Sobre o curso

No dia a dia de equipes de dados, é cada vez mais comum lidar com volumes crescentes de dados, eventos contínuos e múltiplas fontes de ingestão, o que exige compreender como pipelines de dados são estruturados, executados e monitorados em ambientes de nuvem. Para profissionais em início de carreira, entrar nesse contexto sem uma base sólida costuma gerar insegurança e retrabalho.

O curso Google Dataflow – Fundamentos de Pipelines de Dados foi desenvolvido para profissionais que estão dando os primeiros passos em Engenharia de Dados e precisam entender como pipelines de dados funcionam na prática, especialmente nos cenários de batch e streaming no Google Cloud. O foco está na compreensão do papel do Dataflow e do modelo de pipelines do Apache Beam, permitindo que o participante entenda como os dados fluem, são transformados e entregues ao longo do processo.

Ao final do curso, o participante estará mais preparado para ler, executar e acompanhar pipelines existentes, dialogar tecnicamente com equipes mais experientes e evoluir com segurança para responsabilidades maiores, construindo uma base sólida para atuação como Data Engineer Júnior e para os próximos níveis da jornada Dataflow.

Módulos do Curso

Módulo 1 — Introdução ao Google Dataflow

- O papel do Dataflow no ecossistema Google Cloud
- Processamento de dados em batch e streaming
- Casos de uso comuns em engenharia de dados

Módulo 2 — Fundamentos de Apache Beam

- O modelo de programação Beam
- Pipelines, transforms e PCollections
- Conceitos básicos de paralelismo

Módulo 3 — Pipelines Batch no Dataflow

- Estrutura de pipelines em batch
- Leitura e escrita de dados
- Execução e acompanhamento

Módulo 4 — Pipelines de Streaming

- Diferença entre batch e streaming
- Eventos, janelas e noções de tempo
- Primeiros exemplos de pipelines contínuos

Módulo 5 — Monitoramento e Operação Básica

- Execução de jobs no Dataflow
- Logs e métricas
- Cuidados iniciais com custos

O que você vai aprender?

- Compreensão do papel do Google Dataflow na arquitetura de dados
- Entendimento do modelo de pipelines do Apache Beam
- Construção e execução de pipelines simples em batch
- Compreensão dos conceitos básicos de streaming
- Uso inicial do Dataflow Console para monitoramento
- Base técnica para evolução nos níveis Intermediário e Avançado

Pra quem é o curso?

Este curso é indicado para:

- Profissionais iniciantes em Engenharia de Dados
- Analistas de dados em transição para pipelines de dados
- Desenvolvedores iniciando em Data Engineering
- Pessoas que desejam ingressar na trilha de certificação Google Data Engineer

Pré-requisitos

- Noções básicas de programação (Python ou Java)
- Conhecimento introdutório de dados e arquitetura de dados
- Familiaridade básica com Google Cloud ou computação em nuvem

rox