Capítulo 1

O que é engenharia de dados?

Fontes de dados → Engenharia de dados → Analista de dados, Cientistas de dados, Engenheiro de Ias.

Engenharia de dados é um conjunto de técnicas, etapas, processos e procedimentos que permitem pegar os dados em um formato bruto e entregar em um formato utilizável.

Perfil do engenheiro de dados:

Para trabalhar como Engenheiro de Dados é preciso ter perfil técnico, gostar de computação, dominar o uso de sistemas operacionais (especialmente Linux) e desenvolver habilidades nas técnicas, procedimentos e ferramentas usados em engenharia de dados.

Engenheiro de Dados x Cientista de Dados

O Cientista de Dados desenvolve modelos e realiza análises usando Matemática, Estatística, Programação e Machine Learning para explicar e prever comportamentos complexos, resolvendo problemas das áreas de negócio, no mundo real.

Um Engenheiro de Dados projeta e constrói arquiteturas de dados e pipelines para ingestão, armazenamento, processamento e execução de aplicações de grande escala com Big Data.

Os Engenheiros de Dados geralmente coletam dados de diferentes fontes, transformam os dados em diferentes formatos e, em seguida, entregam os dados analistas de Dados, Cientistas de Dados ou Engenheiros de IA. Essa “entrega” pode se dar por meio de repositórios de dados como Data Warehouse e Data Laques, por meio de APIs de acesso, por meio de containers, por meio de pipelines ou outras opções.