

# Fernando Amaral

Profissional de Dados com 15+ anos de experiência, atuando em projetos para empresas como Apple, HCL Tech, Imagen entre outras

Autor de livros, incluindo Introdução à Ciência de Dados

Criador da plataforma Escola de Inteligência Artificial (eia.ai)

Instrutor na Udemy com mais de 280 mil alunos



The background is a complex digital landscape. It features various mathematical expressions such as  $1+x+y+2a+21+$ ,  $\lim_{h \rightarrow 0} \frac{h}{x-12-y+n}$ ,  $x=0 \cdot x^n$ , and  $(1+x+y+2a) - (3a)$ . There are also binary strings like  $11011001001101101011010$  and data-like patterns of dots and lines. The overall color scheme is dark blue and teal with some red highlights.

# Panorama Atual e Futuro da Engenharia de Dados

Fernando Amaral

# O que é Engenharia de Dados?

---

## Definição Técnica:

Engenharia de dados é mover, transformar e disponibilizar dados entre diferentes ambientes, atendendo requisitos de negócio, compliance, segurança e escalabilidade.

## Definição de Negócio:

Engenharia de Dados é a disciplina que garante que a organização tenha dados confiáveis, acessíveis e prontos para serem usados na tomada de decisão, automação e geração de valor.

---



# Contexto Histórico

---

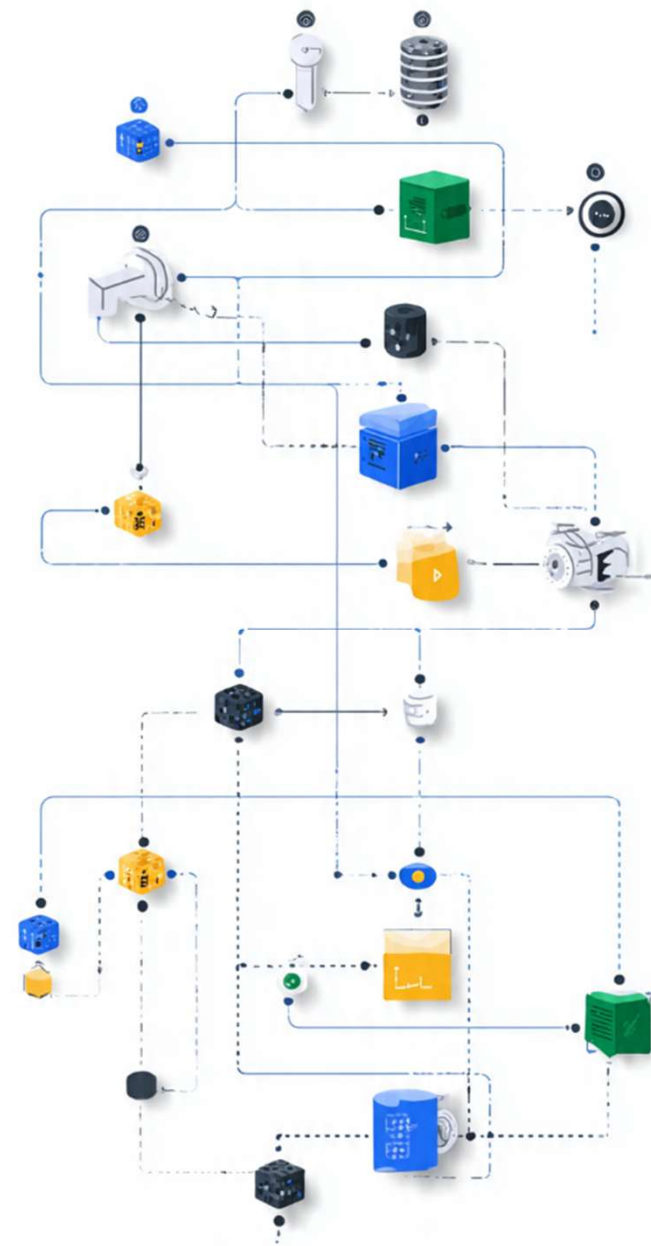
- Antes da Engenharia de Dados: Primeira geração: DBAs e ETL Tradicional (anos 70–2005)
  - Primeira geração: Big Data / Hadoop (2005–2012)
  - Transição: NoSQL + Streaming + Cloud emergente (2012–2016)
  - Segunda geração: Engenharia de Dados Moderna / Cloud (2016–2020)
  - Terceira geração: DataOps, MLOps, Observabilidade (2020–2023)
  - Quarta geração: AI-Native Data Engineering (2023–2025)
- 



# Tipos Principais

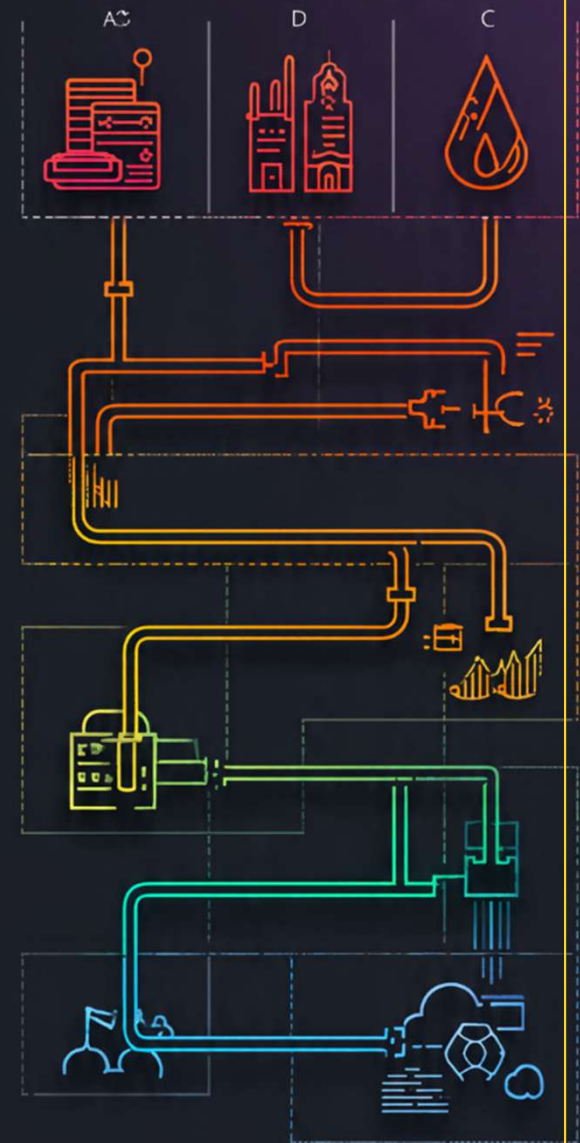
---

- Batch
  - Streaming
- 



# Formatos e Modelos

- Texto e Semi-Estruturado: txt, csv, JSON, XML
- APIS Rest/Sop/GraphQL: Json, XML, CSV
- Colunas e Binários: Parquet, ORC, Avro, Iceberg
- Bancos de Dados Relacionais: Postgres, SQL Server, Oracle
- Bancos de Dados NoSQL: Redis, MongoDB, Neo4j
- Streaming: Kafka, Kinesis
- Dados Proprietários: Excel, SAS, SPSS
- Mídia: Image, Áudio, Vídeo



# Ferramentas

---

- Categorias:
  - Processamento em Batch
  - Processamento de Eventos/Streaming
  - Transformação (ETL/ELT)
  - Orquestração
  - Ingestão
  - Armazenamento
  - Bancos de Dados OLTP e OLAP
  - Governança
  - Observabilidade e Monitoramento
  - Qualidade



# Desafios

---

- Segurança, Compliance e Regulamentação
- Alta Disponibilidade
- Escalabilidade
- Qualidade e Confiabilidade dos Dados
- Complexidade de Integração
- Orquestração e Dependências
- Custos Operacionais e Otimização
- Governança e Catálogo de Dados
- Latência e Performance
- Monitoramento e Observabilidade de Pipelines
- Ambientes Multicloud / Híbridos



## Valor da Informação vs Tempo

Quanto mais tempo demora  
para entregar um dado, menos  
valor ele gera.



# Tempo vs Custo vs Complexidade

---

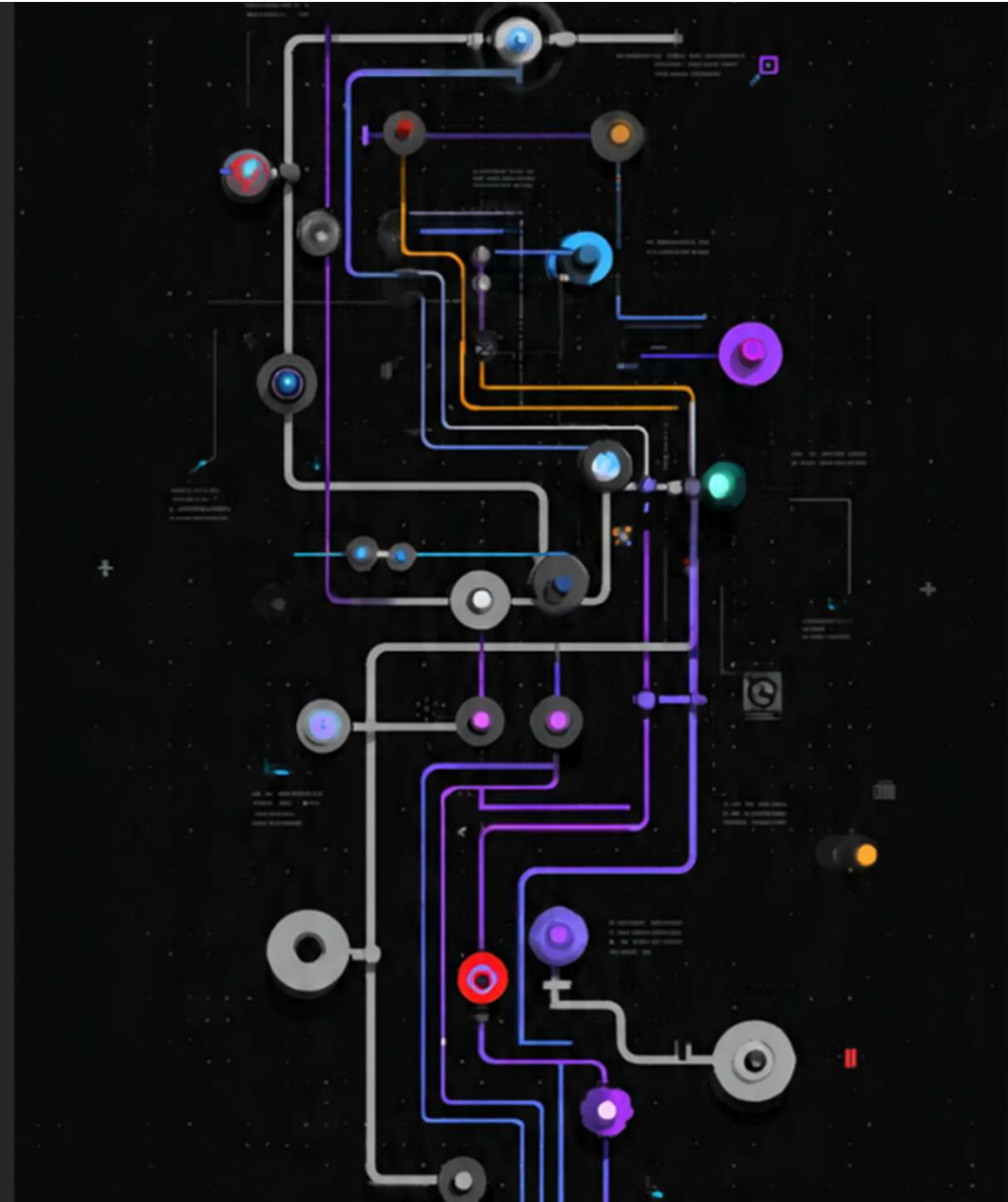
- Reduzir o tempo → aumenta custo e complexidade.
  - Reduzir o custo → aumenta tempo e reduz qualidade.
  - Reduzir complexidade → exige mais tempo ou reduz escopo.
- 



# Disponibilidade vs Custo

---

- 99% (downtime de 3 dias/ano) → barato
  - 99.9% (8h/ano) → caro
  - 99.99% (52 min/ano) → muito caro
  - 99.999% (5 min/ano) → extremamente caro
- 





## Latência vs Custo

---

- Quanto menor a latência desejada, maior o custo computacional.



# Qualidade dos Dados vs Velocidade

Quanto mais rápido você quer entregar o dado, menor a profundidade da validação.



# Escalabilidade vs Complexidade

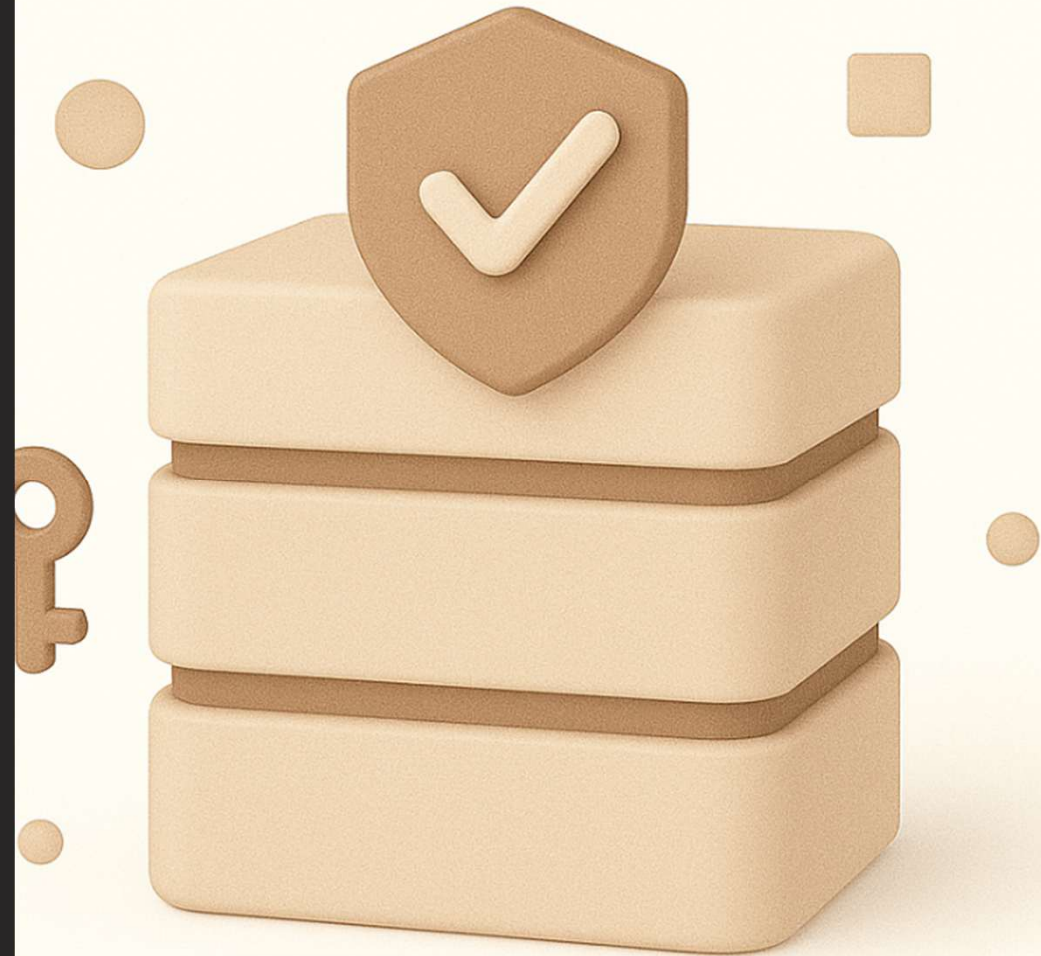
Sistemas muito escaláveis exigem arquiteturas mais complexas.



# Governança vs Agilidade

---

- Quanto mais governança e controle, mais lento é o ciclo de entrega.
- 



# Cases

---

- Crachá Sociométrico
  - Fabrica de Compensados
  - Diagnóstico através de Processamento de Imagens
  - Aplicação Analítica de Câncer de Mama Metastático
  - Construção de DW para fabricante de hardware
- 



Obrigado!

