

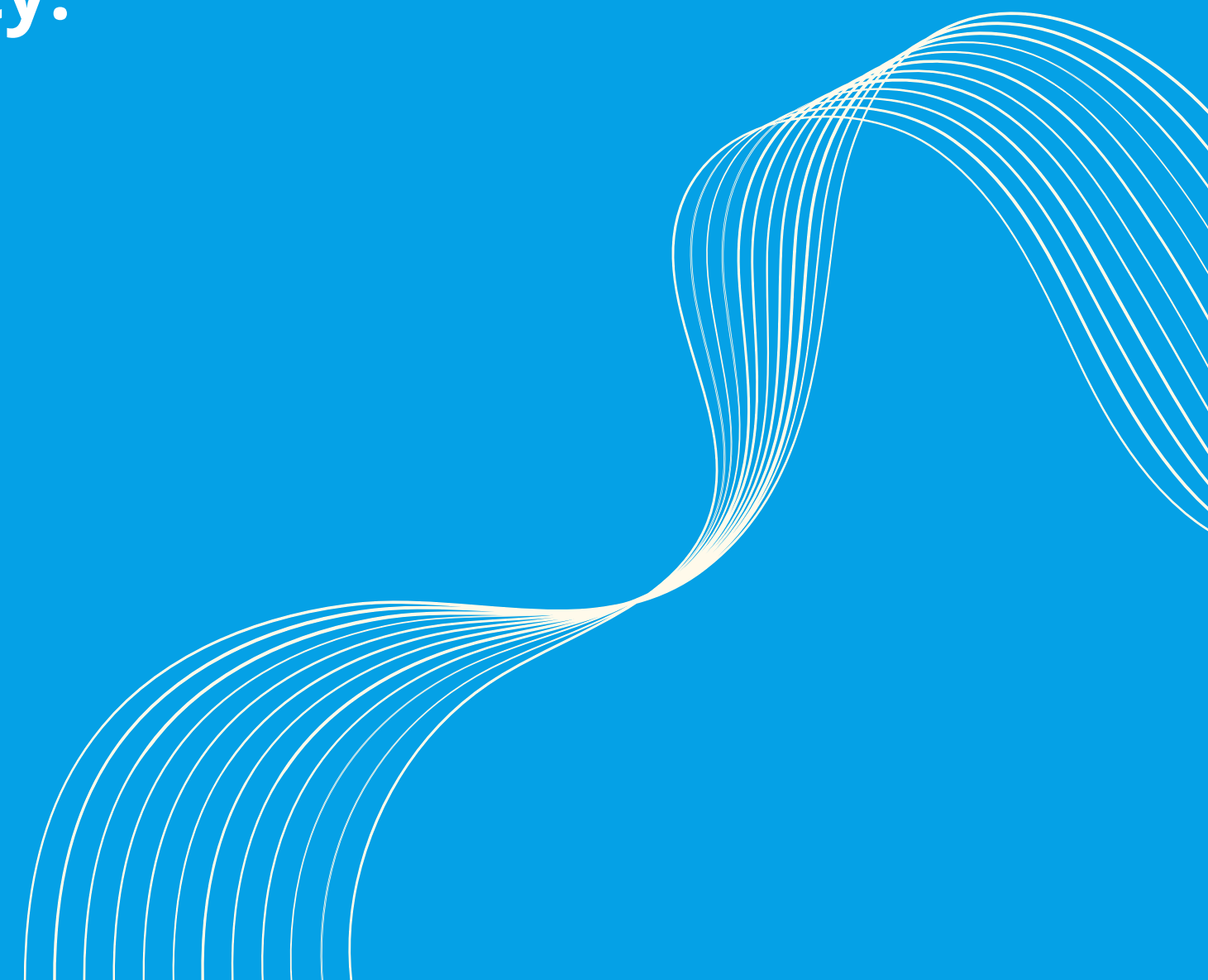
CAPÍTULO 03

DESIGNING MACHINE LEARNING SYSTEMS



Fonte dos dados

- Dados inseridos pelo usuário e dados fornecidos pelo sistema.
- Dados first-party, second-party e third-party.



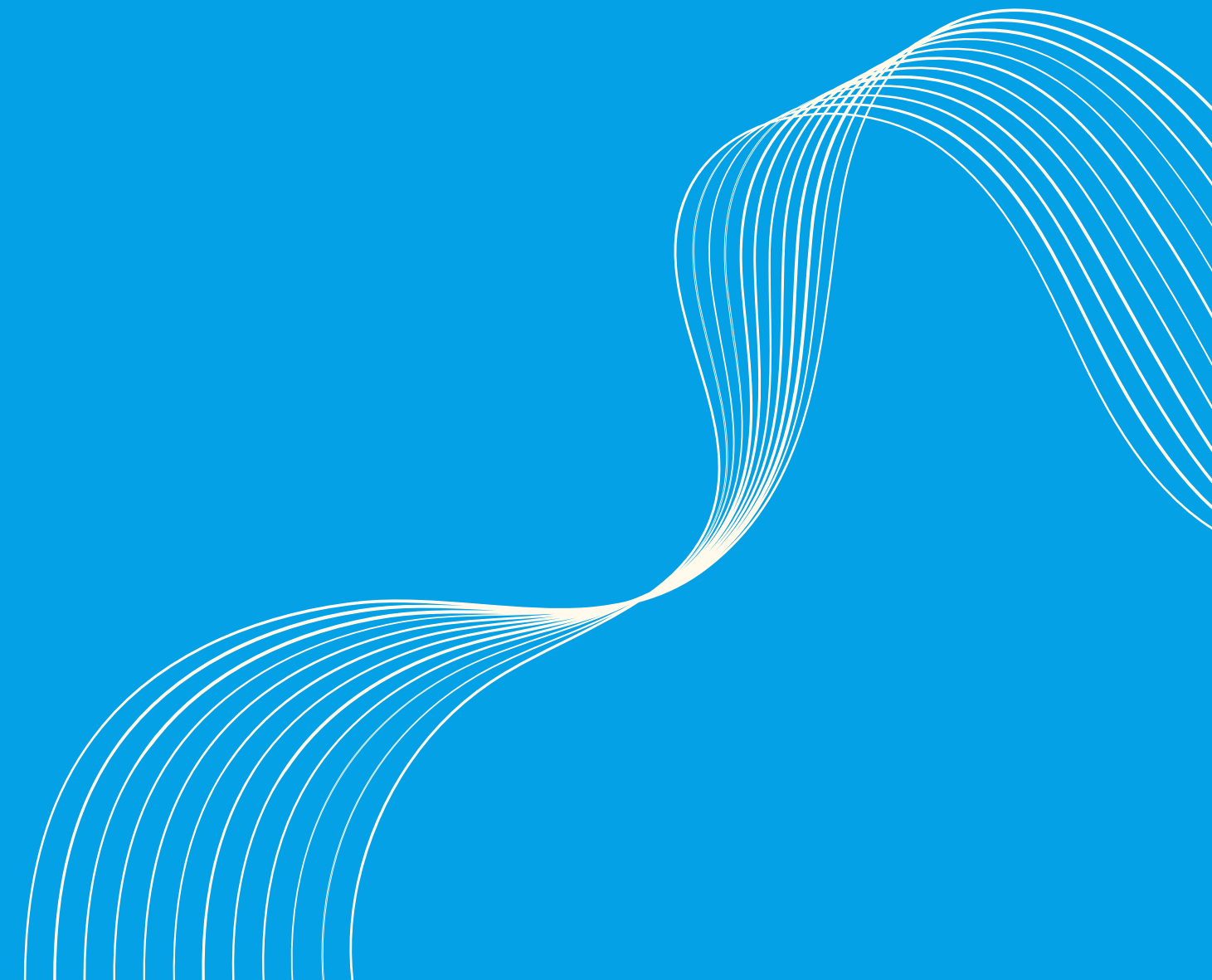
Formato dos dados

Format	Binary/Text	Human-readable?
JSON	Text	Yes
CSV	Text	Yes
Parquet	Binary	No
Avro	Binary primary	No
Protobuf	Binary primary	No
Pickle	Binary	No

Modelos de dados

→ Modelo relacional e NoSQL

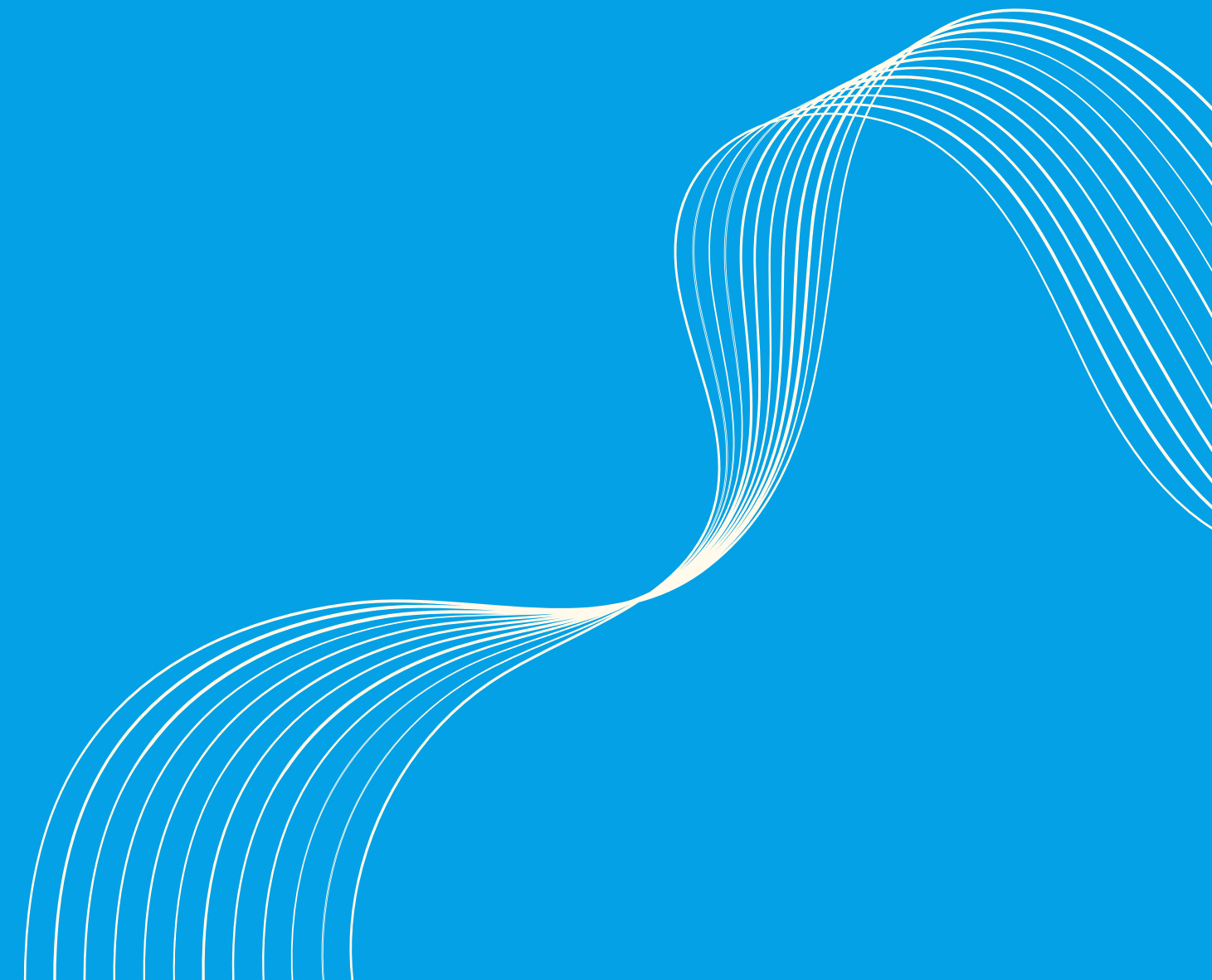
- Orientado a documento
- Orientado a grafo



Mecanismos de armazenamento e processamento

→ Análítico e transacional

→ ETL



Modos de dataflow

- Dados passando por banco de dados
- Dados passando por serviços usando requisições
- Dados passando por um transporte em tempo real

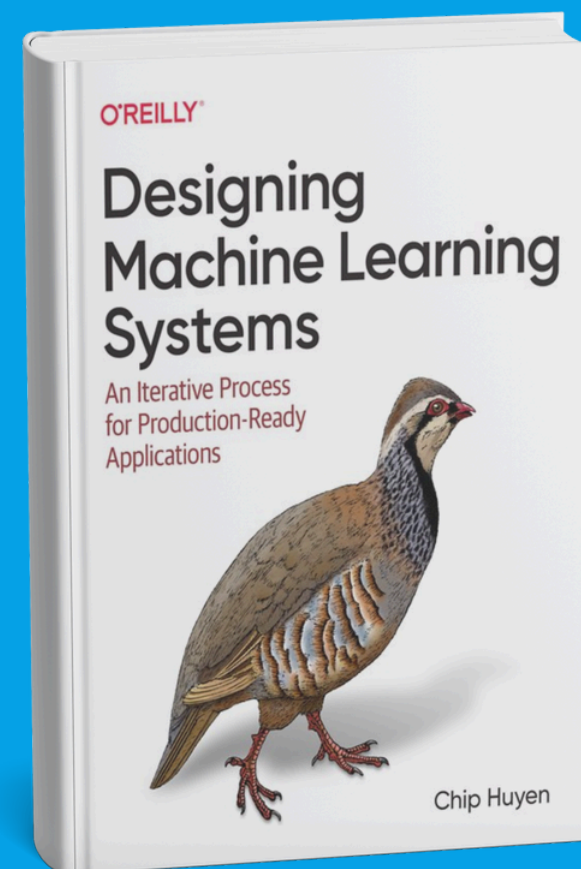


Processamento em lote versus processamento de fluxo



Bibliografia

Designing Machine Learning Systems - Chip Huyen (O'Reilly, 2022)





That's all Folks!