

INTRODUÇÃO À INGESTÃO DE DADOS



INTRODUÇÃO À INGESTÃO DE DADOS

D **DESENVOLVE**
programa de formação
e inclusão de talentos
em tecnologia



 **Grupo
Boticário**



KORU





O FLUXO DE DADOS E SISTEMAS DE ORIGEM



Processos Humanos

Formulários e entrada manual



Sistemas Automatizados

Aplicações e streaming



Armazenamento

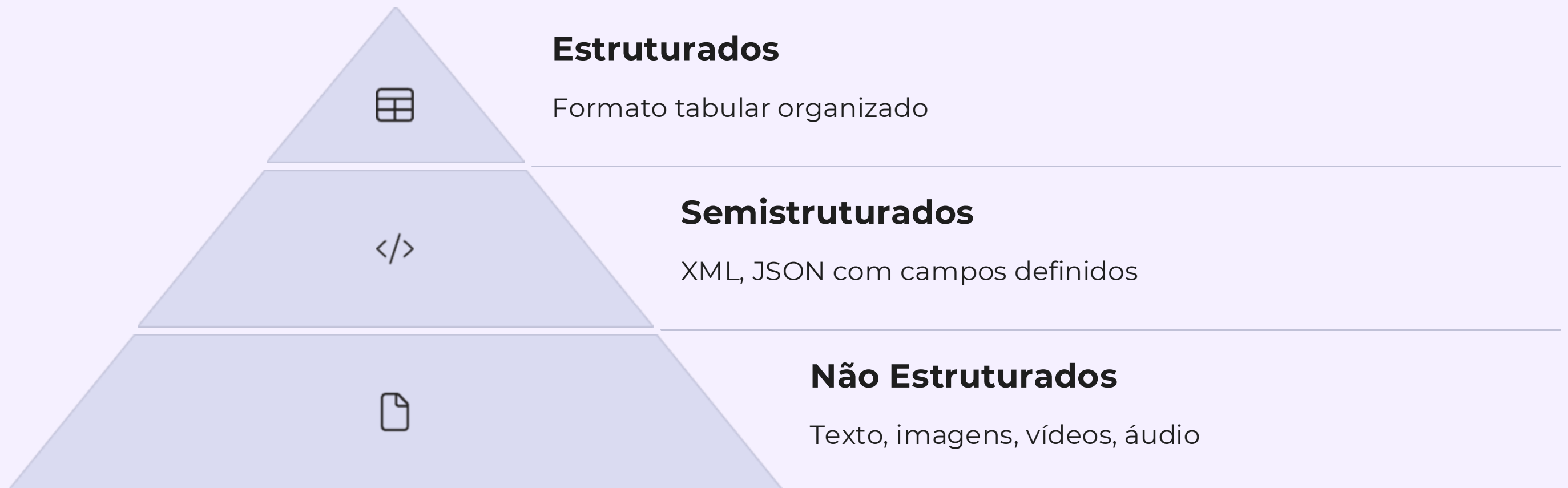
Local, rede ou nuvem



Banco de Dados

Destino final para análise

TIPOS DE DADOS POR ESTRUTURA



FORMATOS DE ARQUIVO COMUNS



CSV

Valores separados por vírgula



JSON

JavaScript Object Notation



XML

eXtensible Markup Language



Parquet

Formato colunar otimizado

IMPORTAÇÃO EM BANCOS SQL

1 Esquema Predefinido

Estrutura deve ser estabelecida antes do carregamento

2 INSERT INTO

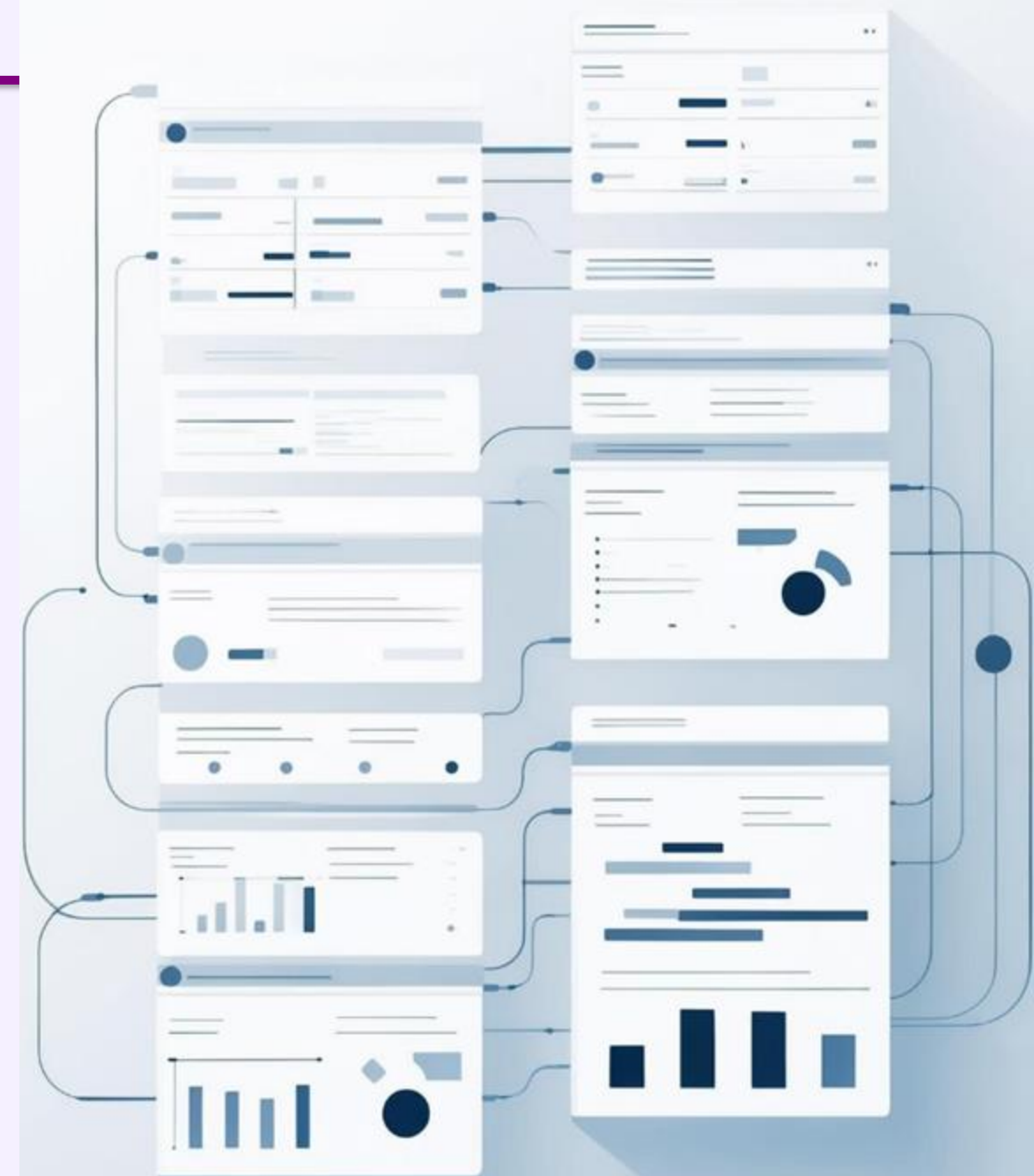
Comando padrão para volumes modestos

3 Bulk Loading

LOAD DATA para grandes volumes

4 Ferramentas Específicas

sqlldr, SQL Developer, APIs dedicadas



PROCESSOS ETL VS ELT

ETL (Extract, Transform, Load)

Transforma dados antes de carregar no destino.

Processamento externo ao banco de dados.

ELT (Extract, Load, Transform)

Carrega primeiro, transforma dentro do banco.

Aproveita poder de processamento do SGBD.

BANCOS NOSQL E ESQUEMAS DINÂMICOS

Key-Value

Armazenamento simples chave-valor



Column Family

Dados organizados em famílias



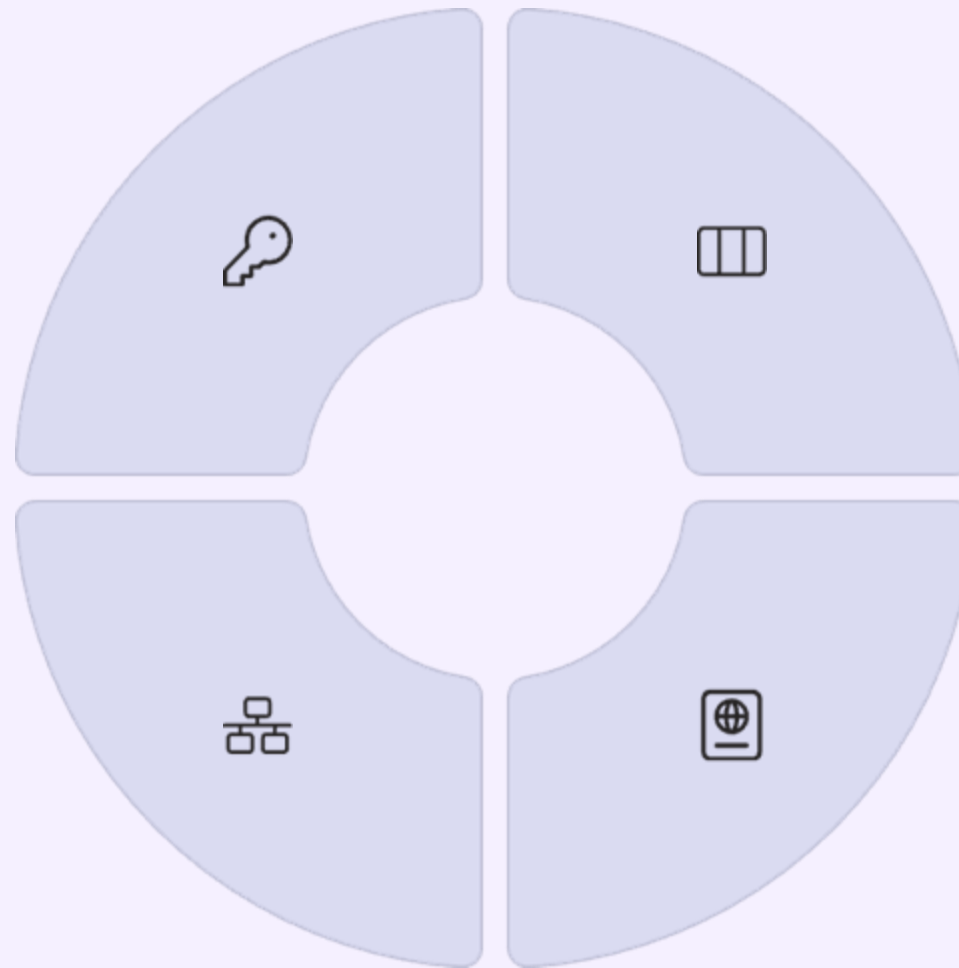
Graph Databases

Relacionamentos em grafos



Document Stores

Documentos JSON/BSON



IMPORTAÇÃO NOSQL VIA APIS

APIs Nativas

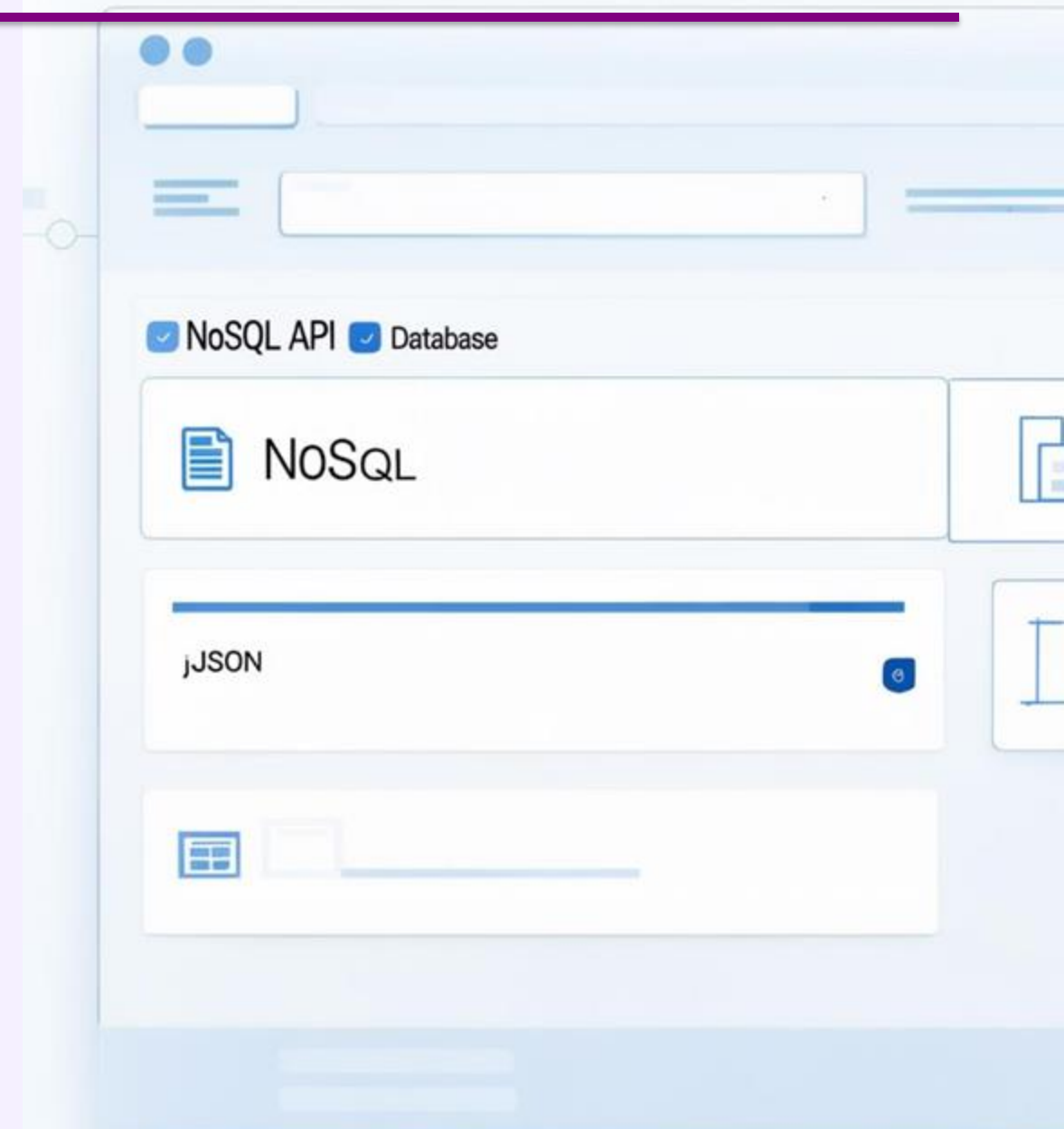
Linguagens específicas como MQL do MongoDB

Carregamento de Documentos

Inserção direta de JSON em coleções

Escalabilidade Horizontal

Sharding para Big Data distribuído



AUTOMAÇÃO DA INGESTÃO



Agendamento

Processos batch ou contínuos



Integração Programática

Python, R, Java para orquestração



Ferramentas Especializadas

Apache Drill para múltiplas fontes



Monitoramento

Alertas e controle de execução

CONSIDERAÇÕES FINAIS

Volume e Velocidade

Escolha da abordagem depende da escala dos dados

Automação Essencial

Fundamental para dados modernos em grande escala

Múltiplas Ferramentas

ETL/ELT, APIs e linguagens de programação

