

PIPELINE BÁSICO DE INGESTÃO

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



 **Grupo
Boticário**



KORU



CONTROLE DE FLUXO DE DADOS



Geração

Dados criados por sistemas de origem



Movimentação

Transferência para ambiente centralizado



Armazenamento

Dados organizados em banco de dados



Análise

Consultas e extração de insights



ETL VS ELT

ETL (Extract, Transform, Load)

Transformação ocorre antes do carregamento no banco de dados.

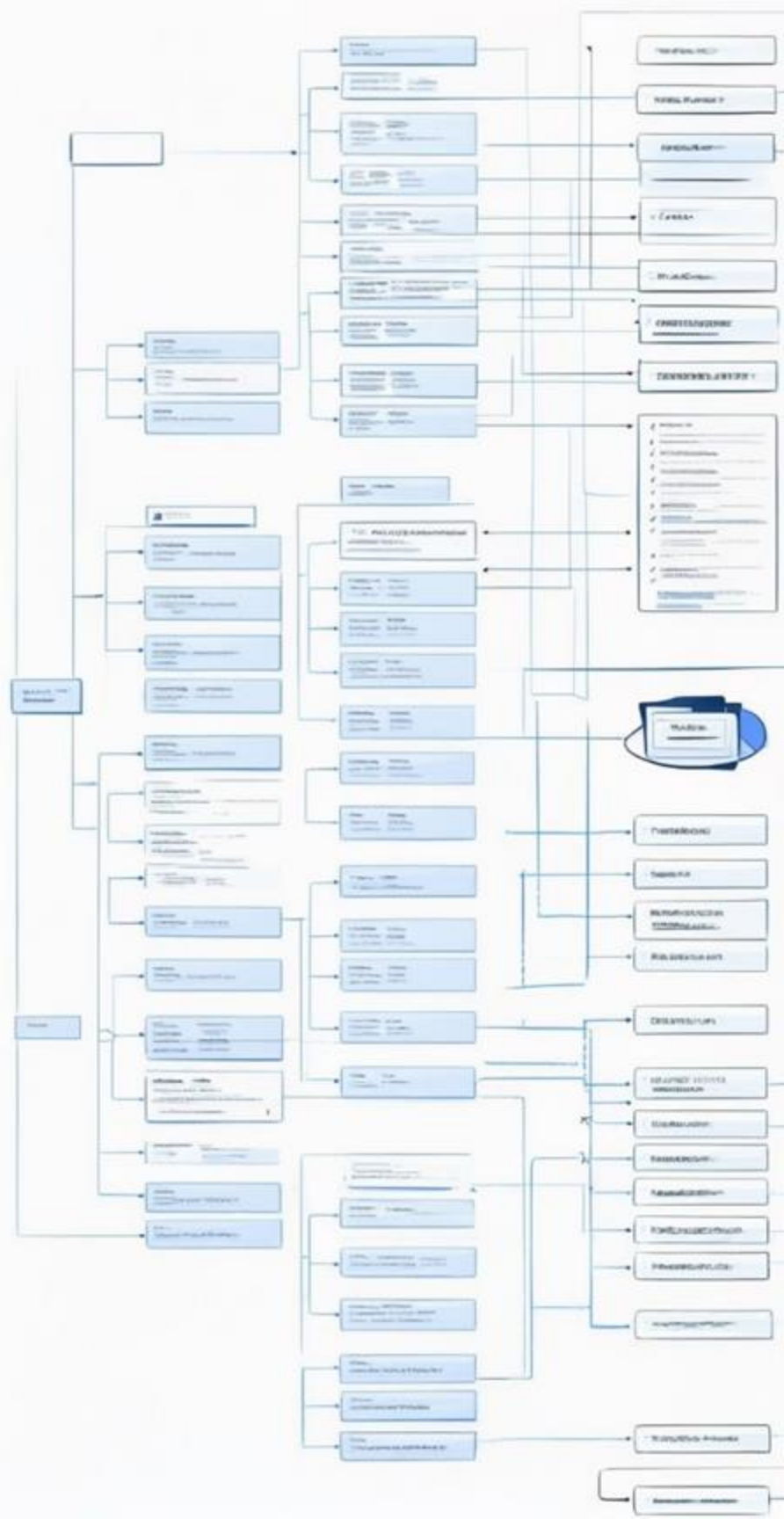
Processo tradicional usado em data warehouses.

ELT (Extract, Load, Transform)

Dados são carregados primeiro, depois transformados.

Aproveita o poder do banco de dados para transformações.

ESTRUTURANDO O ARMAZENAMENTO



Tabelas

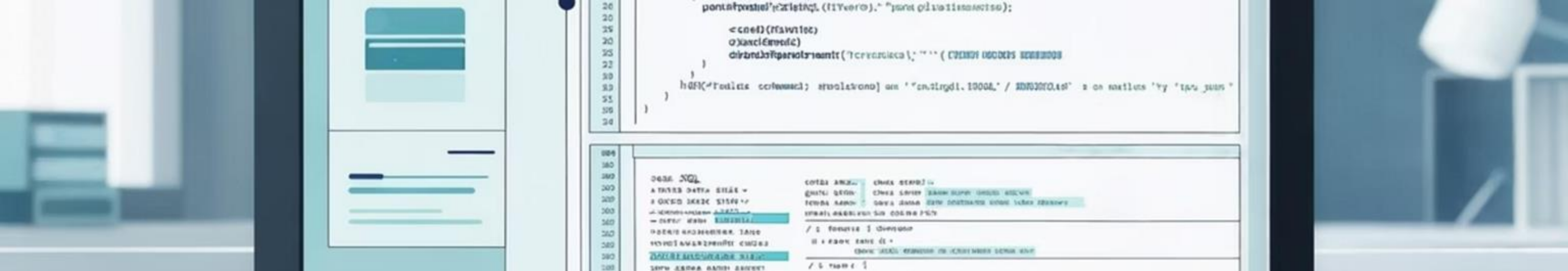
Conjunto de linhas que contém dados, mantido em memória ou armazenamento permanente.

Colunas

Atributos ou campos que definem o tipo de dado permitido em cada parte da tabela.

Modelo de Dados

Descreve as categorias de dados (entidades) e suas relações.



DEFINIÇÃO DO SCHEMA COM DDL

```
CREATE TABLE minha_tabela_simples ( id INTEGER, nome STR), data DATE);
```



Chave Primária

Identifica unicamente cada linha em uma tabela.



Chave Estrangeira

Estabelece relacionamentos entre tabelas.



Normalização

Evita redundância e anomalias no design de tabelas.

An isometric illustration on the left side of the slide depicts a data ingestion process. It features a blue database cylinder with several horizontal segments and small square details. Two large, 3D blue arrows point upwards and to the right, originating from the bottom left and passing behind the database. In the foreground, there is a blue square button with a white circular arrow icon, suggesting a refresh or update action. The background is a light blue grid with faint circuit-like lines and small circles, giving it a technological feel.

CARREGANDO DADOS: INGESTÃO

Comando INSERT

Adiciona linhas individuais à tabela.

Prático para pequenos volumes de dados.

Inserção Múltipla

Permite adicionar várias linhas de uma vez.

Mais eficiente que múltiplos INSERTs individuais.

Carregamento em Massa

Métodos especializados para grandes volumes.

Significativamente mais rápido para conjuntos grandes de dados.

AUTOMATIZANDO A IMPORTAÇÃO

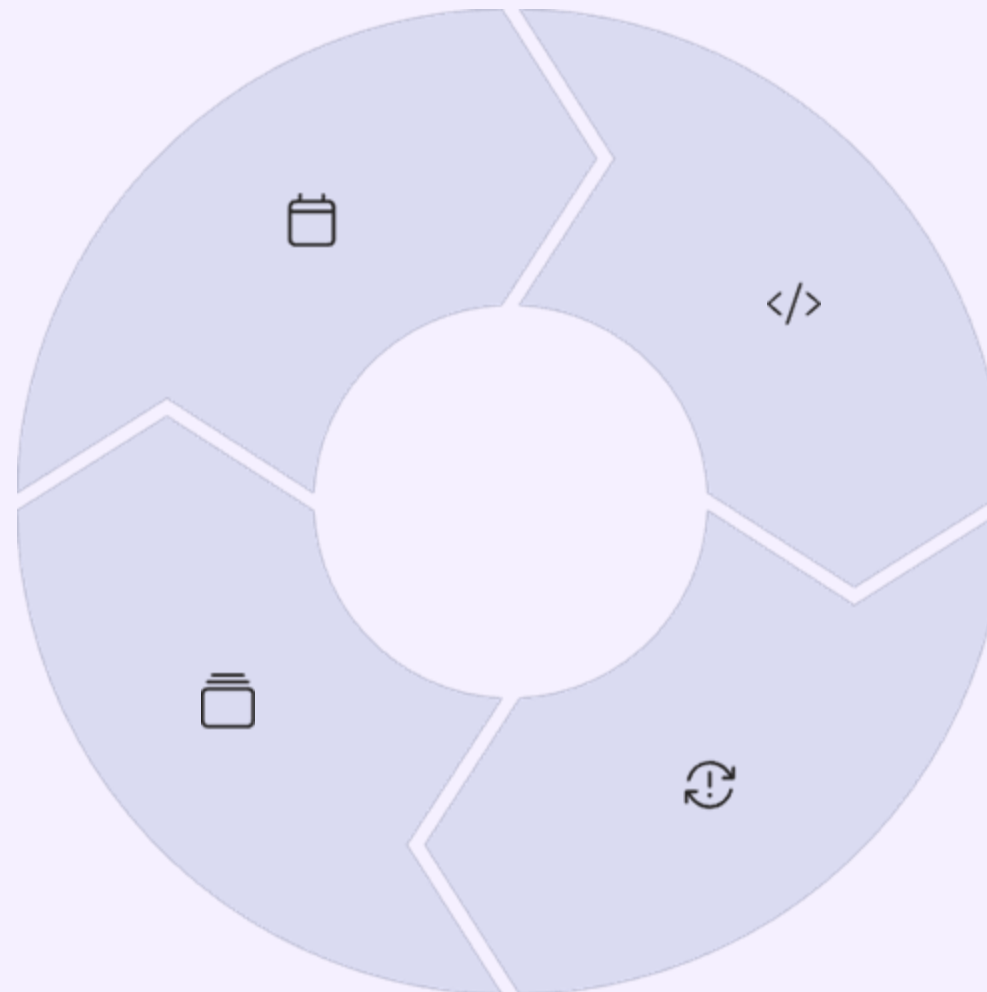
```
LOAD DATA LOCAL INFILE '/caminho/arquivo.csv' INTO TABLE minha_tabela_destino FIELDS TERMINATED BY ',' ENCLOSED BY '"' LINES TERMINATED BY '\n' IGNORE 1 ROWS;
```

Agendamento

Execução programada em intervalos regulares

Histórico

Registro de execuções para auditoria



Orquestração

Gerenciamento via ferramentas ou scripts externos

Monitoramento

Verificação de erros e status de execução

ANALISANDO DADOS COM SQL



SELECT-FROM-WHERE

Estrutura básica de consulta SQL



Funções de Agregação

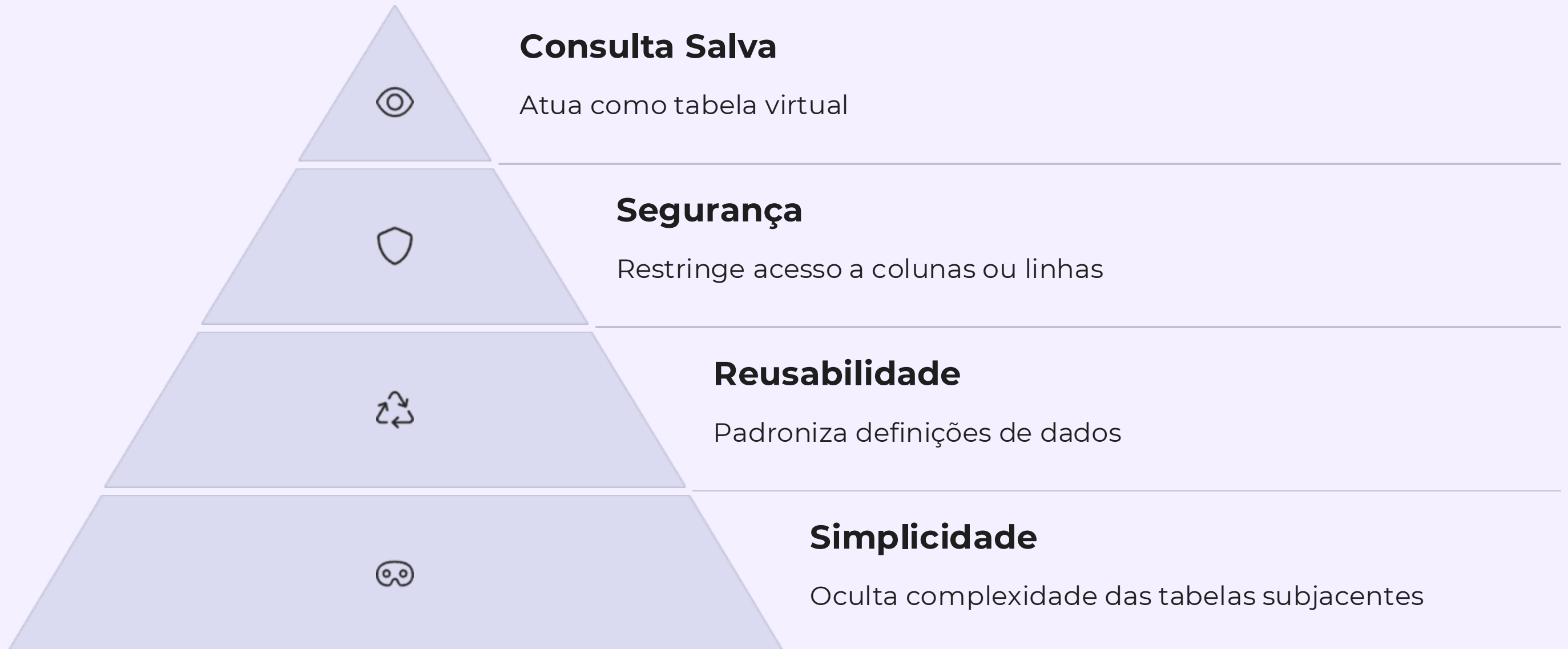
COUNT(), SUM(), AVG(), MIN(), MAX()



GROUP BY

Agrupamento de resultados para análise

VISÕES (VIEWS)



FLUXO COMPLETO DE INGESTÃO E ANÁLISE



Definir Estrutura

CREATE TABLE para estabelecer o schema



Carregar Dados

INSERT ou métodos de bulk loading



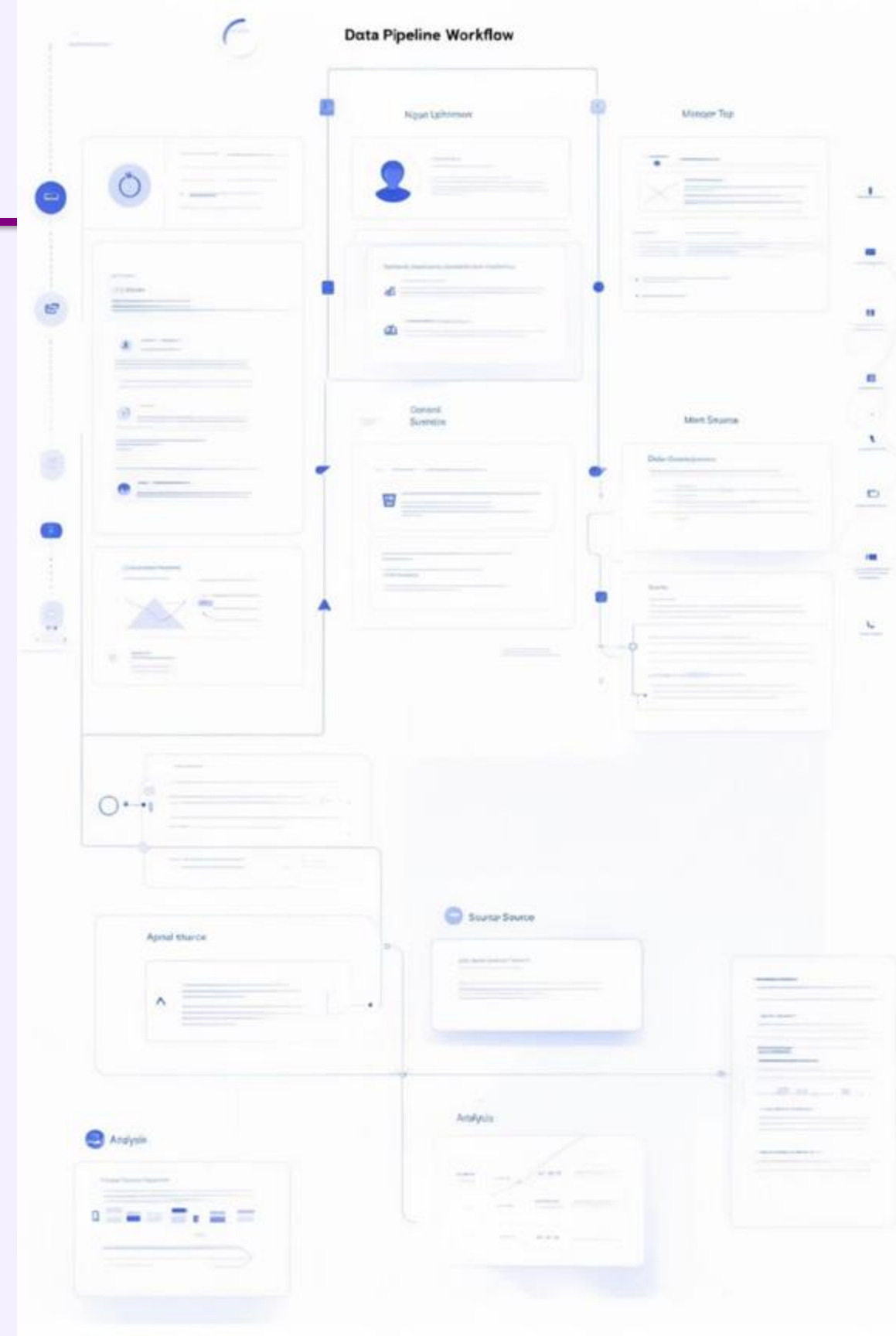
Automatizar

Processos ETL/ELT para atualização contínua



Analisar

SELECT e Views para consulta e insights





**Momento de
avaliar a aula
de hoje!**