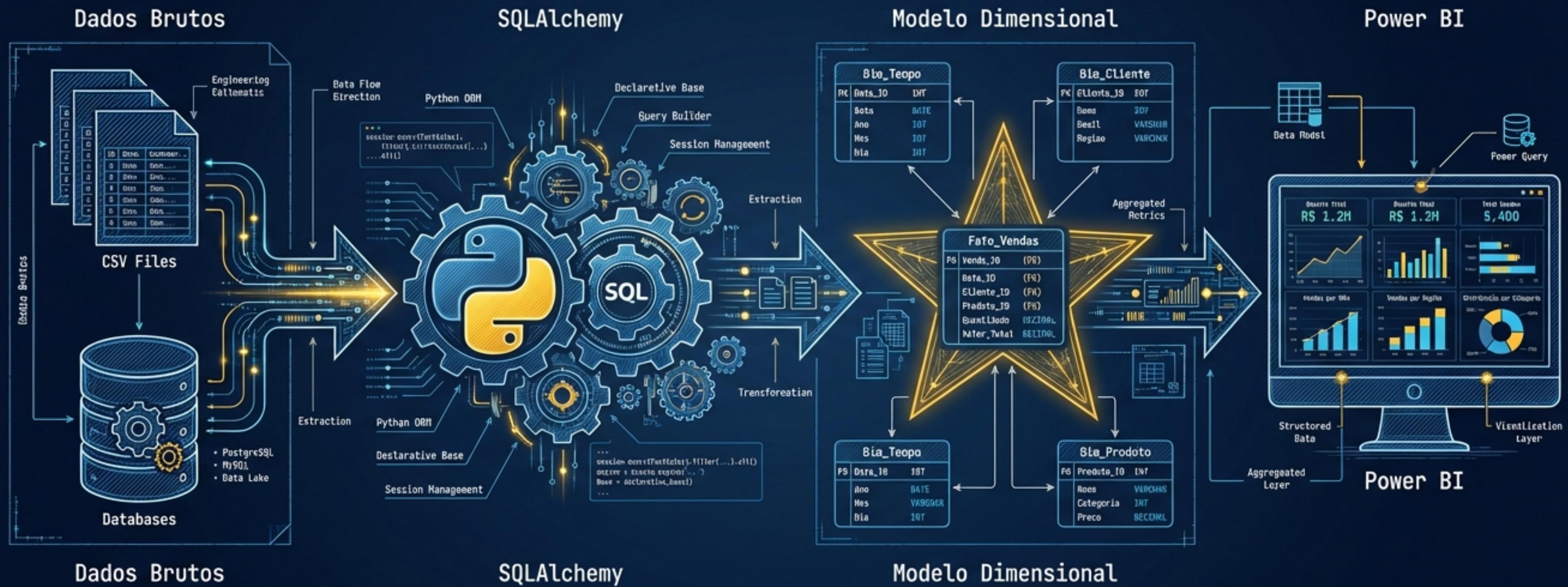


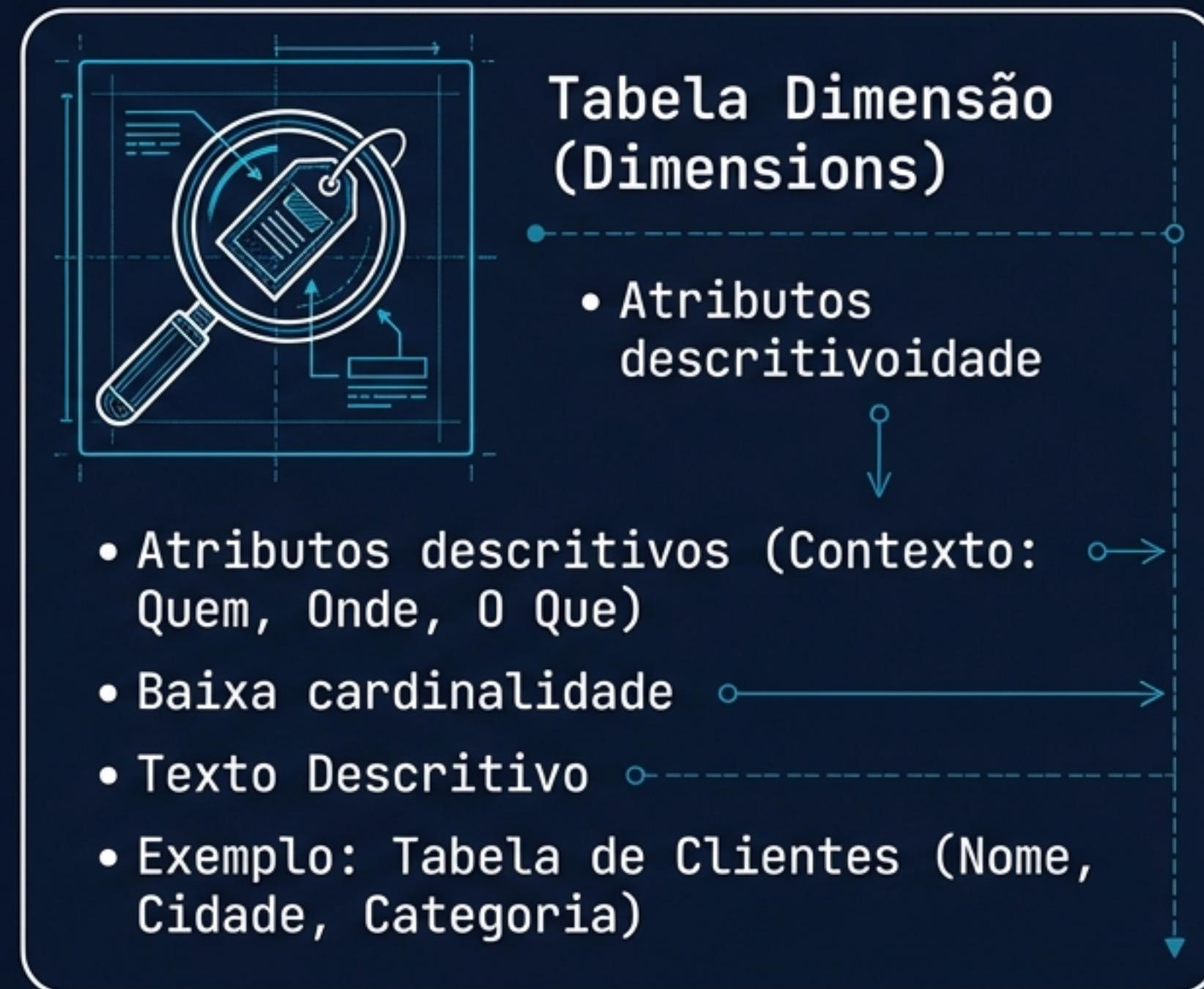
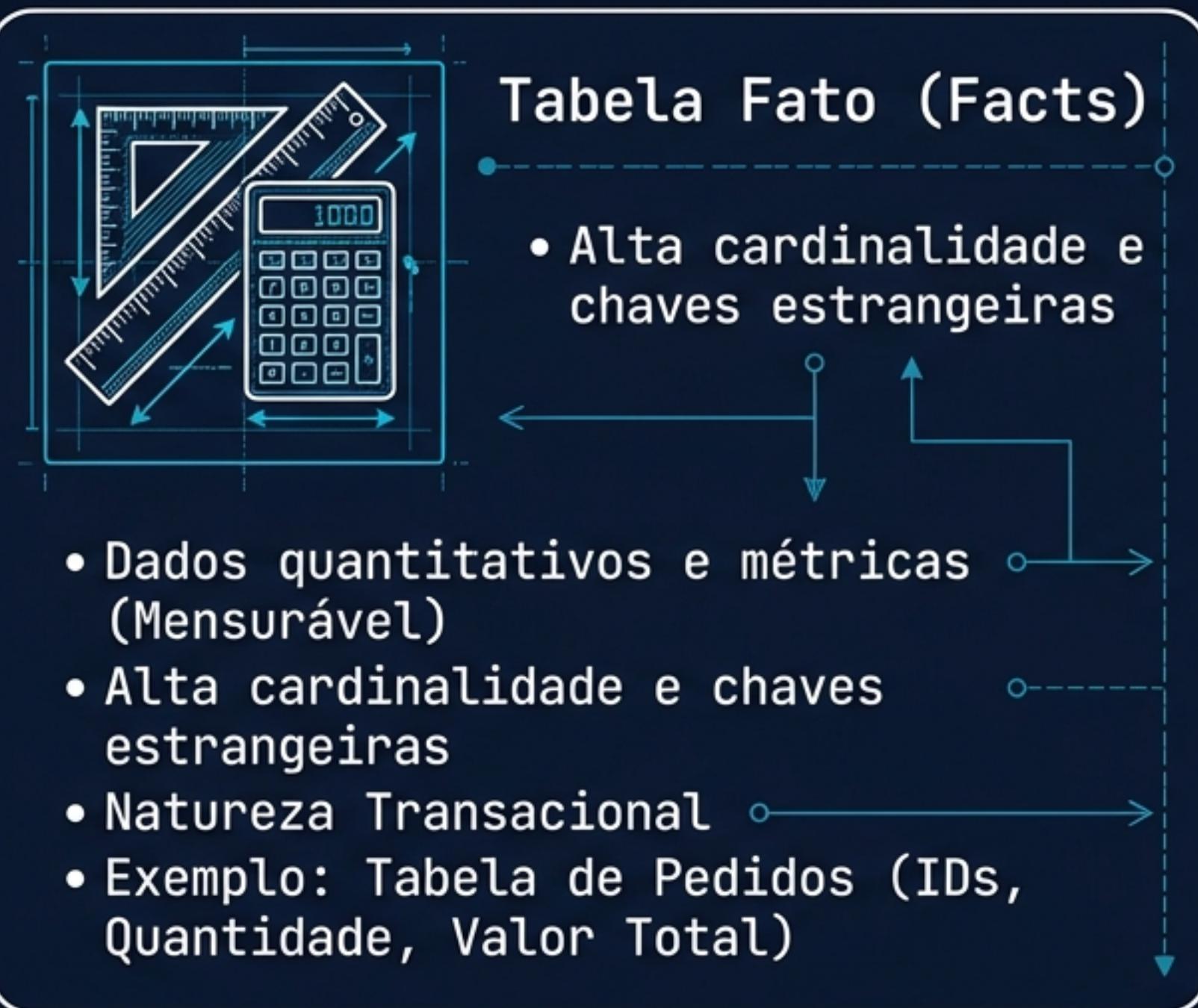
Blueprint de Arquitetura Moderna de Dados

Da Modelagem Dimensional ao Power BI via Python e SQLAlchemy



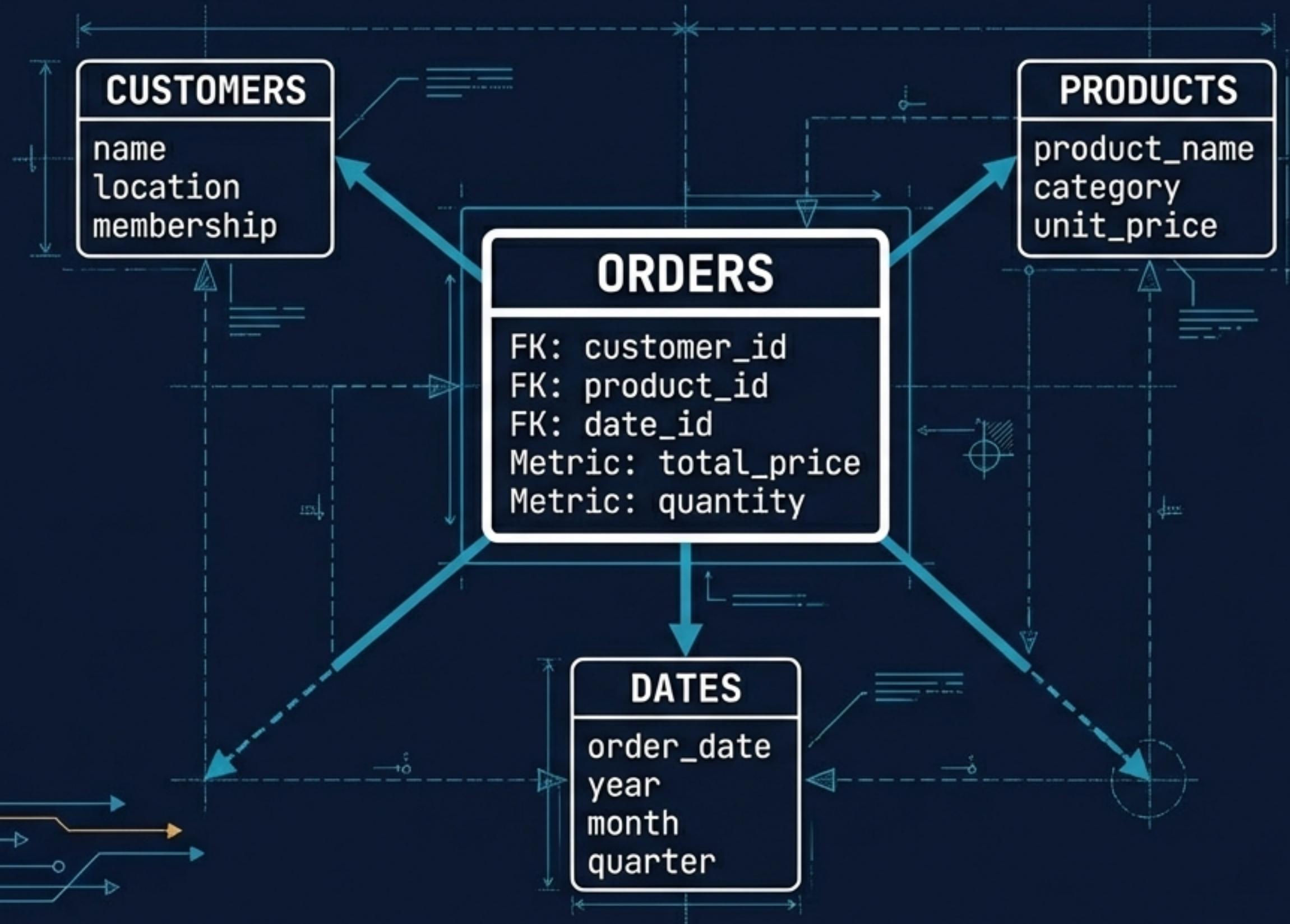
Um guia visual para analistas, engenheiros de dados e desenvolvedores de BI.

O Alicerce do Data Warehousing: Fatos vs. Dimensões



»»» O objetivo é organizar dados dispersos em estruturas otimizadas para consulta (OLAP), não apenas transação (OLTP).

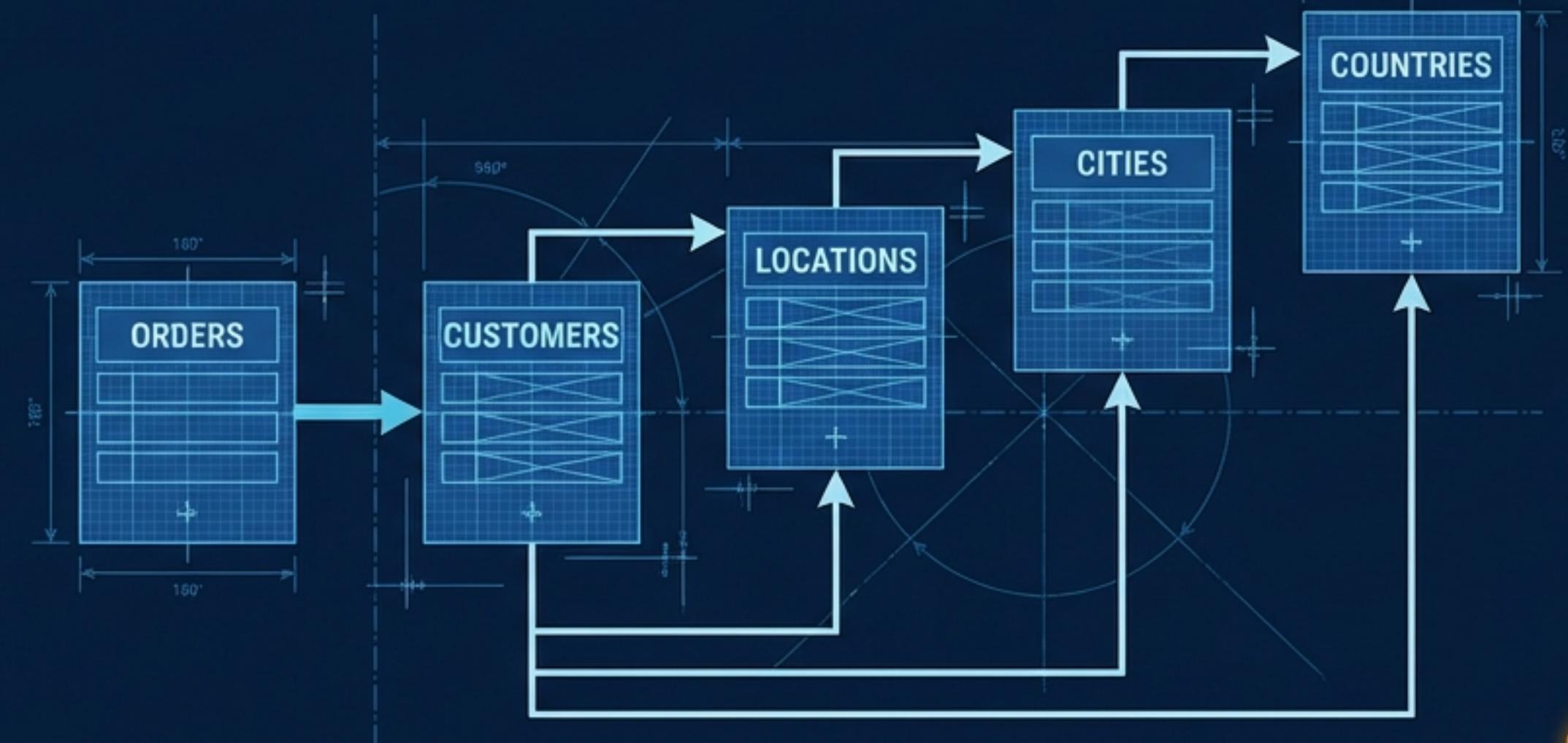
O Padrão Ouro: Star Schema (Esquema Estrela)



- **Estrutura:** Fato central cercada por dimensões desnormalizadas.
- **Performance:** Minimiza joins em consultas SQL.
- **Legibilidade:** Intuitivo para ferramentas de BI.

A Alternativa Normalizada: Snowflake Schema

- **Conceito:** Normalização das dimensões em sub-tabelas.
- **Vantagem:** Redução de redundância e eficiência de armazenamento.
- **Desvantagem:** Alta complexidade de joins e manutenção difícil.



Ideal para dados com hierarquias profundas onde o espaço de armazenamento é uma preocupação crítica.

Batalha dos Esquemas: Star vs. Snowflake no Power BI

Star Schema

- **Performance:** ★★★★★ (Leituras rápidas)
- **Engine Power BI:** Otimizado para VertiPaq
- **DAX:** Simples e Direto

Snowflake Schema

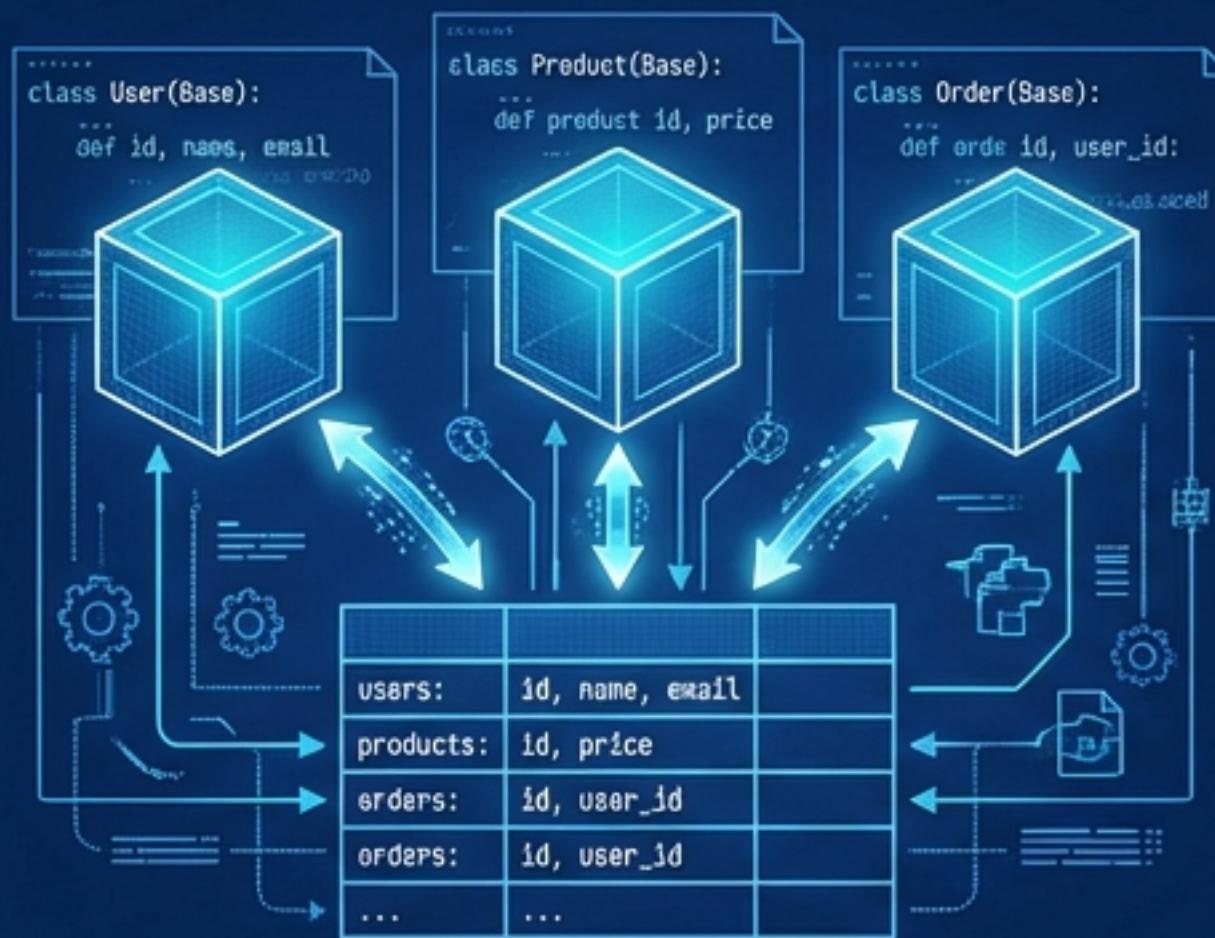
- **Performance:** ★★☆ (Overhead de múltiplos joins)
- **Engine Power BI:** Custo de processamento maior
- **DAX:** Complexo (Filtros em cadeia)

Insight de Especialista

Para a maioria das soluções em Power BI, o Star Schema é preferido. O ganho de performance e a simplicidade do DAX superam a economia de armazenamento do Snowflake.

SQLAlchemy: A Ponte entre Python e SQL

ORM (Object Relational Mapper)



- Mapeamento Objeto-Relacional
- Classes Python <-> Tabelas
- Foco: Lógica de Negócios

Core (SQL Expression Language)



- Controle granular de Schema
- Abstração de SQL puro
- Foco: Analytics e Migrações

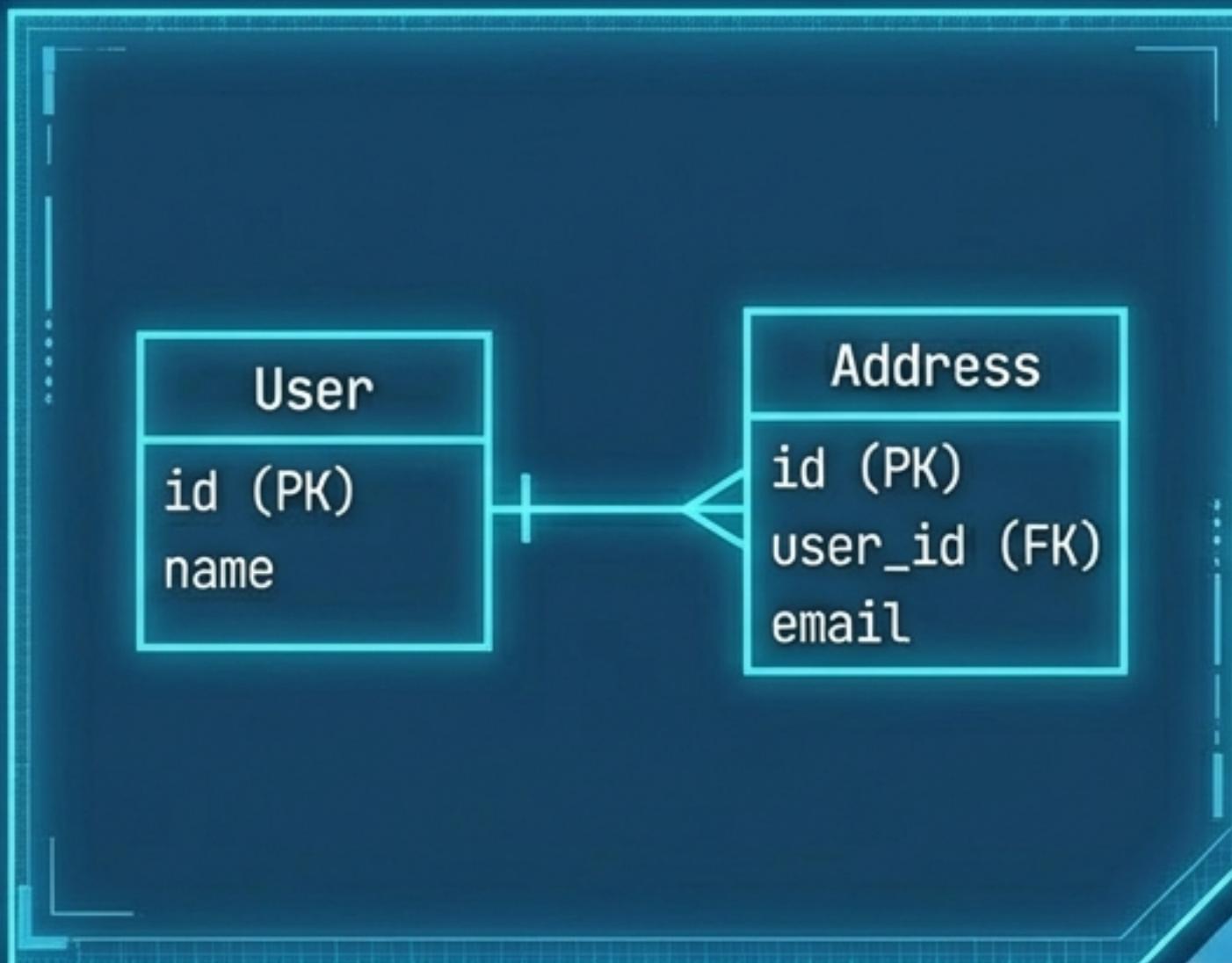
O toolkit padrão para abstração de banco de dados sem perder o poder do SQL.

Arquitetura Técnica: Do Engine ao Banco de Dados



```
engine = create_engine('postgresql+psycopg2://user:pass@host/db')
```

Definindo o Modelo Dimensional com ORM Declarativo



```
class User(Base):
    __tablename__ = 'users'
    id = Column(Integer, primary_key=True)
    name = Column(String)

    addresses = relationship('Address', back_populates='user')

class Address(Base):
    __tablename__ = 'addresses'
    id = Column(Integer, primary_key=True)
    user_id = Column(Integer, ForeignKey('users.id'))
    user = relationship('User', back_populates='addresses')
```

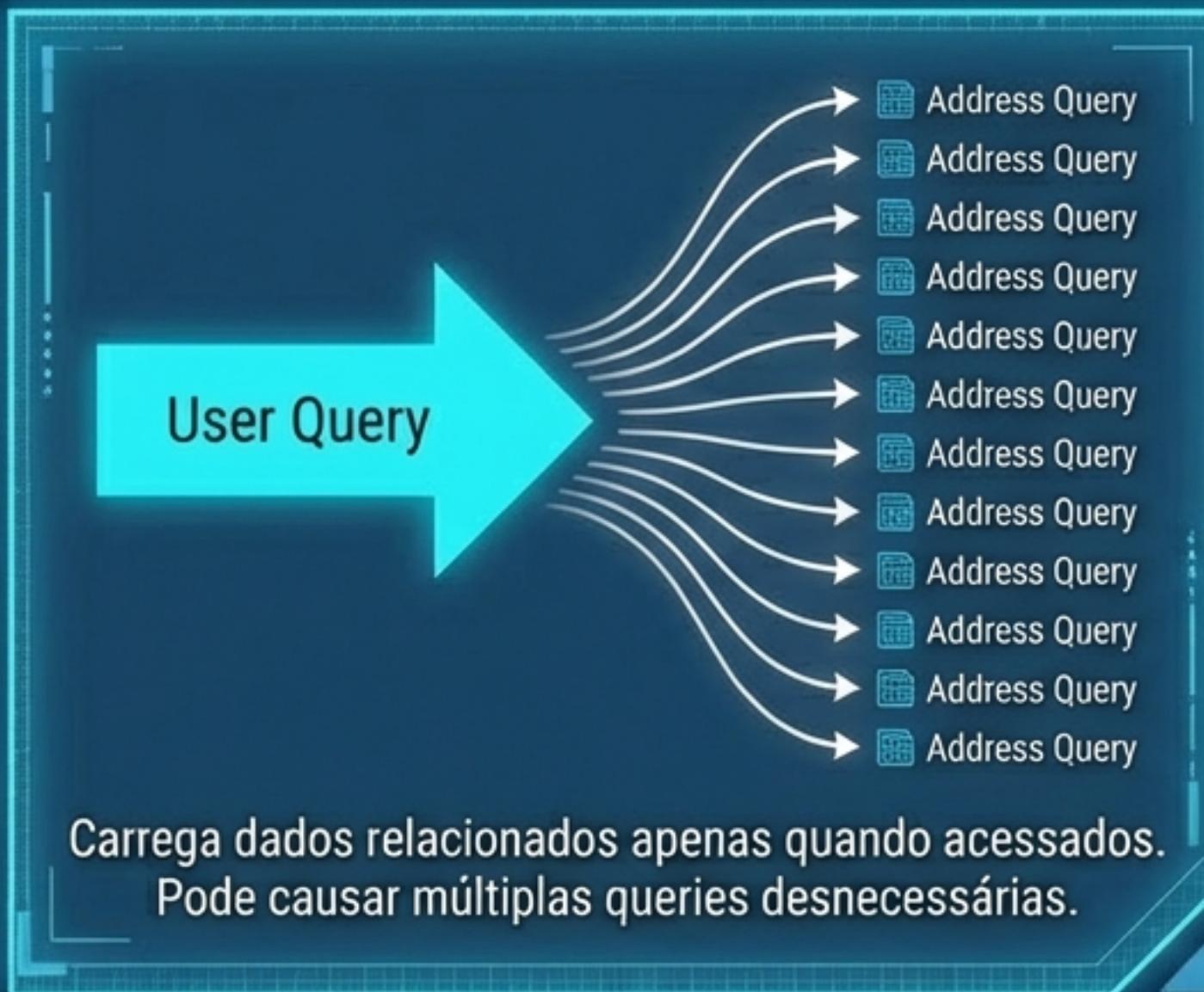
O sistema declarativo une a classe Python e a tabela SQL.

Gerenciamento de Sessão e Operações CRUD

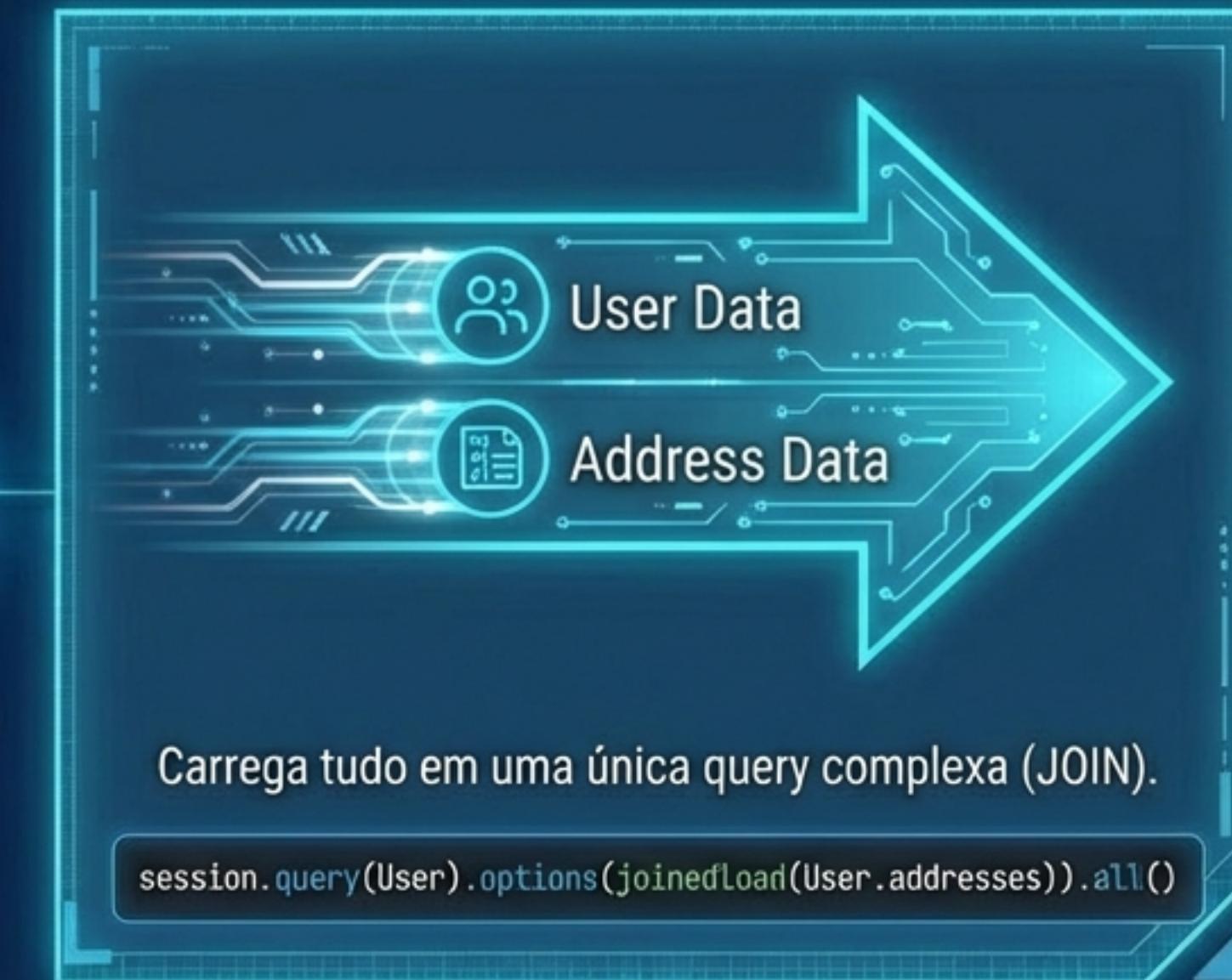


Otimização de Performance: Loading Strategies

Lazy Loading (Risco N+1)



Eager Loading (Optimized)



⚠️ Use Bulk Operations para cargas massivas de ETL.

Potencializando o Power BI com Python



Automação de ETL

Scripts para limpeza e transformação (Pandas) que superam o Power Query nativo.

Analytics Avançado

Uso de bibliotecas estatísticas (NumPy, SciPy) diretamente no fluxo de dados.

Visualização Customizada

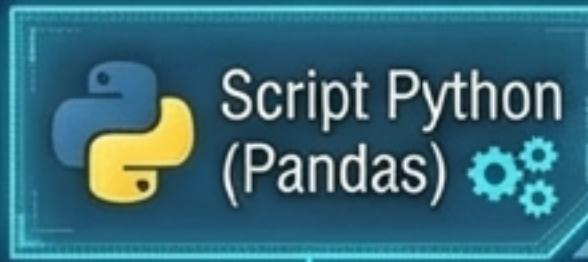
Criação de gráficos complexos (Violin, Heatmap) via Matplotlib e Seaborn.

Configuração: File > Options > Python scripting

Python como Motor de ETL no Power Query



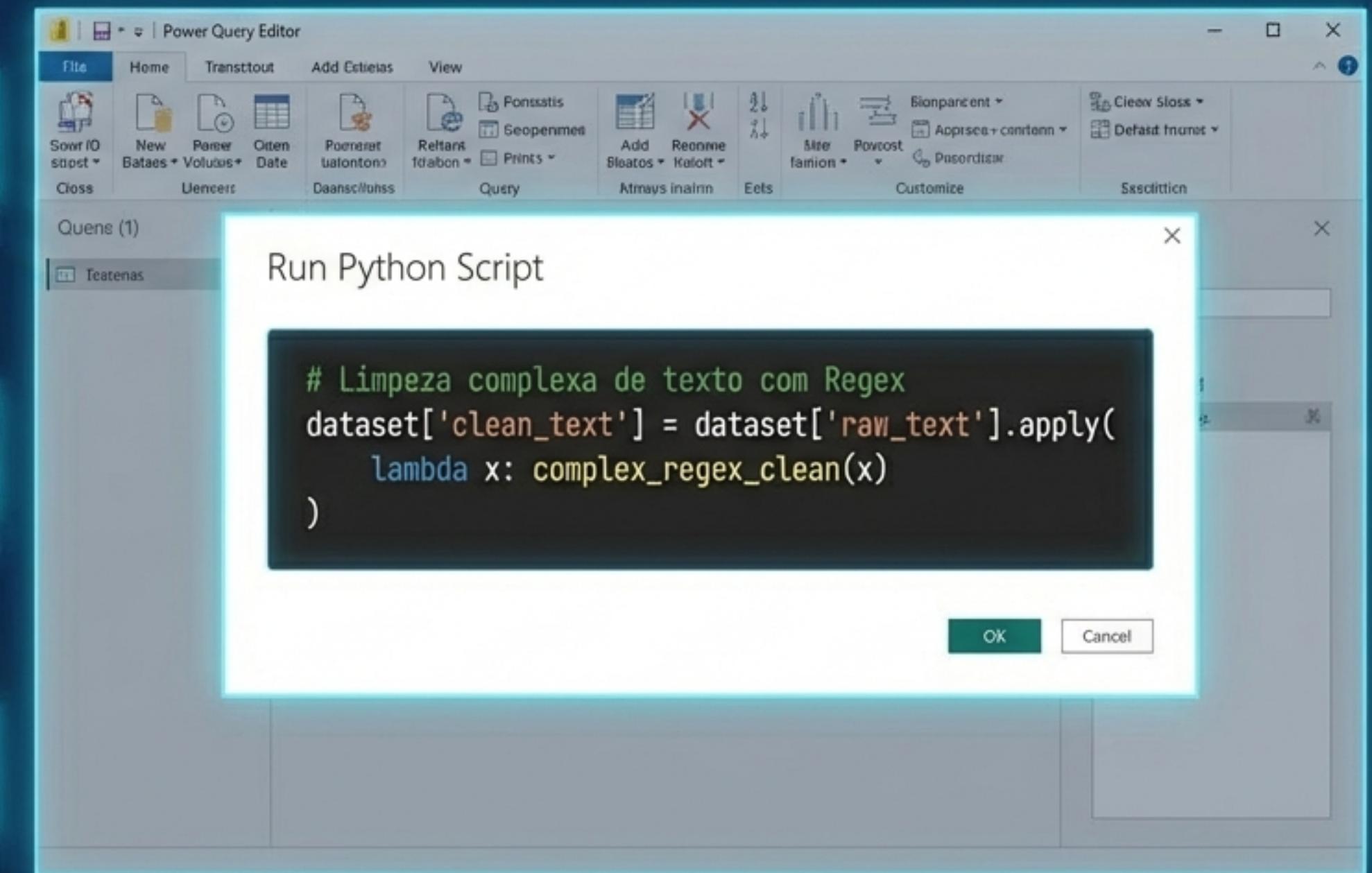
Dados Brutos



Script Python
(Pandas)



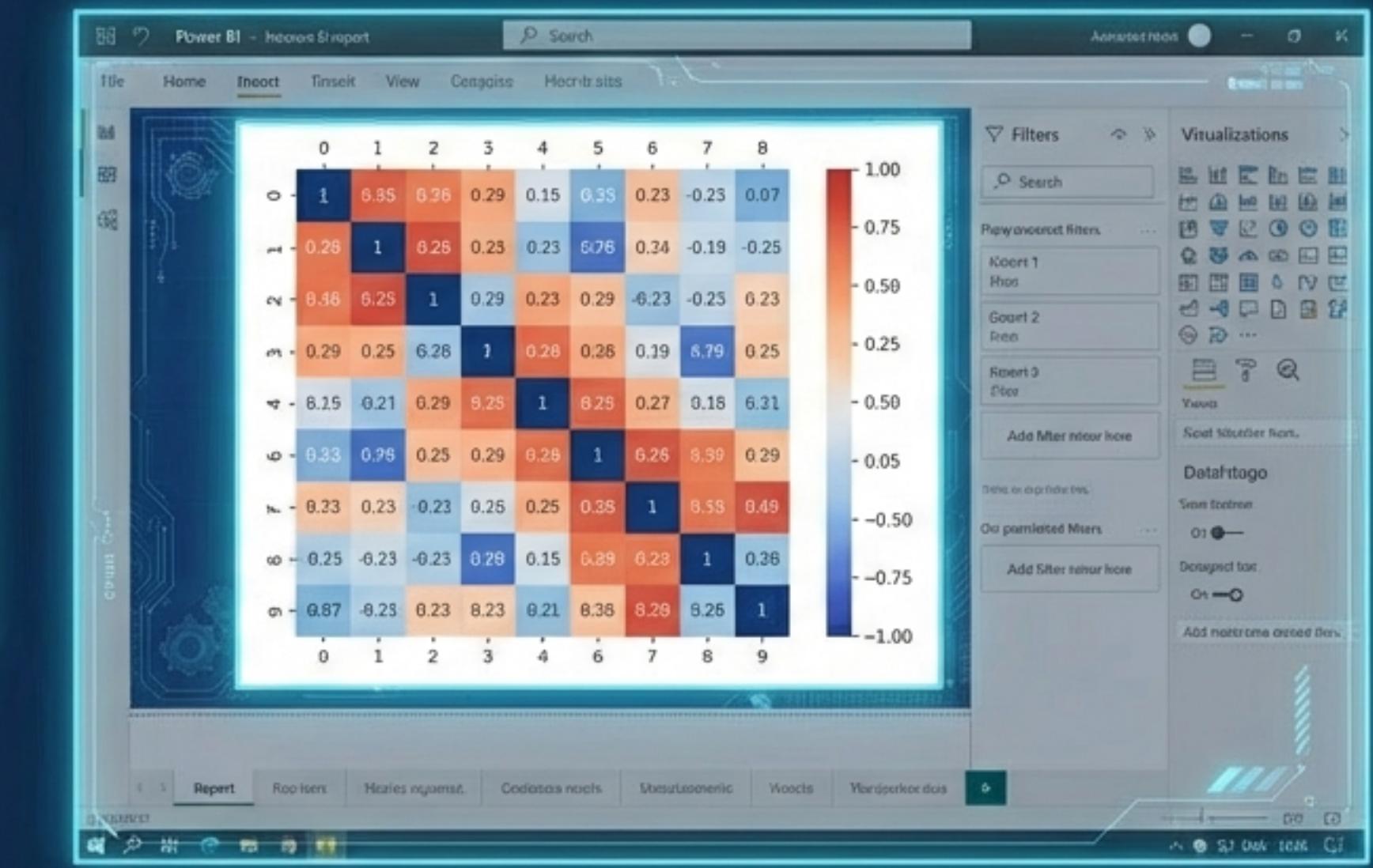
DataFrame
Limpo



Ideal para
pré-processamento,
Regex avançado e
imputação de
dados faltantes.

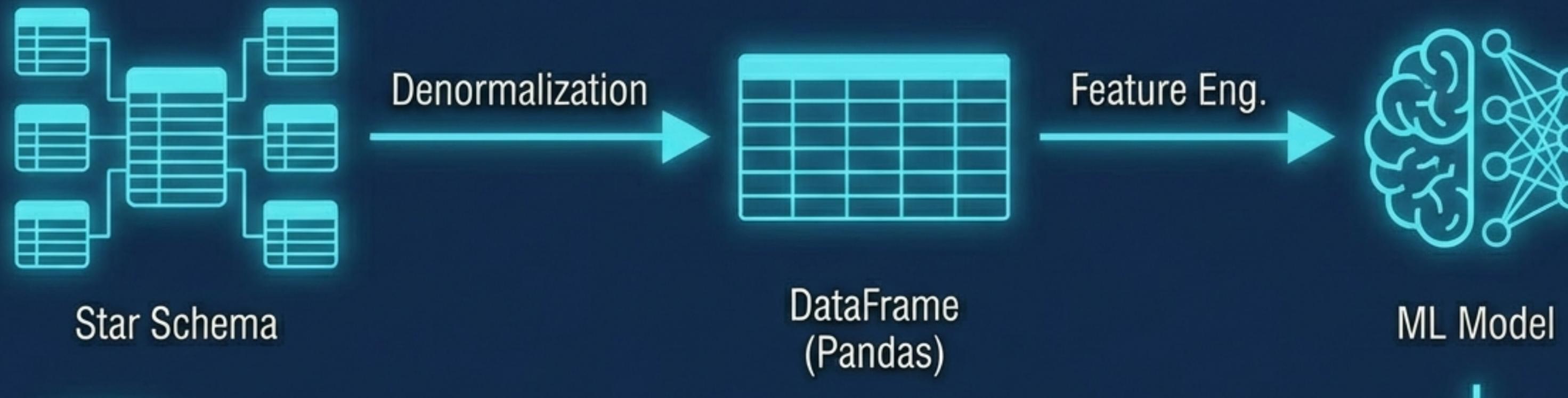
Visualização Customizada: Matplotlib & Seaborn

```
import matplotlib.pyplot as plt  
import seaborn as sns  
  
sns.heatmap(dataset.corr(), annot=True)  
plt.show()
```



Visuais estáticos, mas com profundidade estatística superior aos nativos.

Modelagem Dimensional para Machine Learning



Dados estruturados do Star Schema reduzem o tempo de limpeza e alimentam modelos preditivos consistentes.

Melhores Práticas e Recursos Essenciais

Checklist de Arquitetura



- Comece com Star Schema; migre para Snowflake apenas se necessário.
- Use SQLAlchemy Sessions com context managers ('with').
- Documente scripts Python no Power BI (caixas pretas).
- Prefira Power Query nativo para tarefas simples (Performance).

Biblioteca de Referência

- The Data Warehouse Toolkit (Ralph Kimball)
- SQLAlchemy 2.0 Documentation
- Python for Data Analysis (Wes McKinney)

A arquitetura moderna é a orquestração harmônica entre SQL, Python e BI.