**Cenário**  
Sua empresa vende produtos on-line e mantém registros em sistemas diferentes: um sistema de vendas, um CRM de clientes e uma plataforma de eventos web. Esses dados apresentam duplicidades, formatos inconsistentes e informações faltantes.

**Objetivo do Desafio**  
Construir um pipeline que integre esses três conjuntos de dados, aplique um modelo de dados normalizado e execute processos de limpeza robustos, de modo a gerar uma base única pronta para análises.

**1. Arquivos Fornecidos**

* **orders.csv**
  + Colunas: order\_id, customer\_id, product\_code, order\_date (formatos mistos), amount (valores monetários com “,” e “.”).
* **customers.json**
  + Atributos: id, full\_name, email, registration\_date, state (siglas e nomes por extenso).
* **web\_events.csv**
  + Colunas: event\_id, user\_email, event\_type, event\_timestamp, metadata (JSON string).

Você pode disponibilizar esses arquivos num repositório Git ou como anexos, conforme sua preferência.

**2. Tarefas a Realizar**

1. **Integração de Dados**
   * Identifique e implemente as chaves de junção adequadas entre os três conjuntos (e.g. customer\_id ↔ id ↔ user\_email).
   * Defina o fluxo (ETL/ELT) que mova os dados de origem para o ambiente final (qualquer tecnologia à sua escolha: SQL, Python, ferramenta ETL etc.).
2. **Modelo e Normalização**
   * Proponha um esquema relacional em, no mínimo, terceira forma normal (3FN).
   * Documente o diagrama de entidades (tabelas, chaves primárias e estrangeiras).
3. **Limpeza e Padronização**
   * Trate duplicatas (por exemplo, clientes com e-mails repetidos ou nomes com variações).
   * Uniformize formatos de data e valores monetários.
   * Padronize siglas e nomes de estados.
   * Extraia campos úteis de colunas JSON (ex.: em metadata de eventos).
   * Registre quaisquer suposições ou regras de negócio aplicadas.
4. **Entrega de Resultados**
   * **Código ou Notebook** (Python, SQL, R ou ferramenta de sua escolha).
   * **Diagrama do Modelo de Dados** (pode ser em PDF, imagem ou texto formatado).
   * **Relatório Técnico** (breve – até 2 páginas) que inclua:
     + Descrição do pipeline e ferramentas utilizadas.
     + Principais desafios e como foram resolvidos.
     + Exemplos de antes/depois da limpeza (ex.: valores corrigidos).

**3. Critérios de Avaliação**

* **Qualidade da Integração:** correção das junções e robustez do fluxo.
* **Aderência ao Modelo Normalizado:** clareza e completude do diagrama.
* **Efetividade da Limpeza:** tratamento de inconsistências e duplicidades.
* **Documentação e Legibilidade:** código organizado, comentários e relatório claro.
* **Criatividade e Justificativa:** escolhas técnicas bem fundamentadas sem “receita de bolo”.

**Observação:**  
O objetivo é avaliar seu raciocínio e autonomia — **não** há um caminho único. Você pode usar SQL puro, notebooks Python, Ferramentas ETL comerciais ou open-source. O importante é apresentar uma solução coerente, escalável e bem documentada.