

Teste Engenheiro de Dados

Sua empresa, uma Fintech que realiza empréstimos para produtores rurais, trabalha com diferentes tipos de títulos de crédito (CCBs(Cédulas de Crédito Bancario), CPRs(Cédula de Produto Rural) e Notas Promissórias) e processos de pagamento (boletos gerados via API e registros de pagamentos em arquivos posicionais).

Desafio 1

Descrição:

Os dados das CCBs estão inicialmente registrados em um banco MongoDB e precisam ser transferidos para um banco Postgres para outras operações. Forneça uma solução que realize essa transferência de forma periódica e destaque o método utilizado.

Schema:

```
{
  "_id": "K9NBx1d00WcMEpOBKmwg",
  "clientId": "qbxL3XKAL7M0ev9rvs94RbPYAma",
  "name": "FERNANDO PEIXOTO CAMILO",
  "taxId": "05633022456",
  "type": "PF",
  "birthDate": "1980-12-08",
  "maritalStatus": "casado",
  "identity": {
    "number": "24.094.218",
    "issuer": "PCMG",
    "issuerState": "MG"
  },
  "birthPlace": "Santa Maria do Suacui",
  "address": {
    "zipCode": "39775-000",
    "state": "MG",
    "city": "José Raydan",
    "address": "CORREGO BOA VISTA",
    "neighbourhood": "Rural",
    "number": "0"
  },
  "pep": false,
  "structure": "FIDC - Nagro",
  "state": "MG",
  "motherName": "Irani Camilo Peixoto",
  "gender": "masculino",
  "contacts": {
    "email": "fernandocamilo654@gmail.com",
```

```
    "phone": "(11) 95502-9939"
  },
  "averageMonthlyIncome": 10788,
  "bankData": {
    "branchNumber": 4103,
    "accountNumber": 8079,
    "bankNumber": "756"
  },
  "guarantors": null,
  "operation": {
    "VencimentoPrimeiraParcela": "2021-12-07",
    "Carencia": null,
    "FinanIOF": true,
    "Conta": "1100000000",
    "ValorBruto": "15448.41",
    "ValorDoSeguro": "0",
    "ValorDaCAD": "0",
    "ValorDoIOF": "448.41",
    "ValorLiquido": "15000.0",
    "TaxaDeJuros": "2.8178",
    "TaxaDeJurosAnual": "39.58",
    "CET": "3.22",
    "CET_ANUAL": "46.27",
    "DataDeLiberacao": "2021-10-07",
    "DataDeVencimentoInicial": "2021-12-07",
    "DataDeVencimentoFinal": "2022-11-07",
    "NumeroDeParcelas": 12,
    "parcelas": {
      "PrevisaoDeParcela": [
        {
          "NumeroDaParcela": "0",
          "DataDeVencimento": "2021-11-07T00:00:00",
          "ValorDaAmortizacao": "0",
          "ValorDaCorrecao": "0",
          "ValorDoJuros": "449.82",
          "ValorDoSeguro": "0",
          "ValorTaxaBancaria": "0",
          "ValorDaPrestacao": "0",
          "ValorDoSaldoAnterior": "15448.41",
          "ValorDoJurosCapitalizados": "0",
          "ValorDoSaldoAtual": "15898.23"
        },
        {
          "NumeroDaParcela": "1",
          "DataDeVencimento": "2021-12-07T00:00:00",
          "ValorDaAmortizacao": "1131.87",
          "ValorDaCorrecao": "0",
          "ValorDoJuros": "447.98",
          "ValorDoSeguro": "0",
          "ValorTaxaBancaria": "0",
```

```
"ValorDaPrestacao": "1579.85",
"ValorDoSaldoAnterior": "15898.23",
"ValorDoJurosCapitalizados": "0",
"ValorDoSaldoAtual": "14766.36"
},
{
  "NumeroDaParcela": "2",
  "DataDeVencimento": "2022-01-07T00:00:00",
  "ValorDaAmortizacao": "1163.76",
  "ValorDaCorrecao": "0",
  "ValorDoJuros": "416.09",
  "ValorDoSeguro": "0",
  "ValorTaxaBancaria": "0",
  "ValorDaPrestacao": "1579.85",
  "ValorDoSaldoAnterior": "14766.36",
  "ValorDoJurosCapitalizados": "0",
  "ValorDoSaldoAtual": "13602.6"
},
{
  "NumeroDaParcela": "3",
  "DataDeVencimento": "2022-02-07T00:00:00",
  "ValorDaAmortizacao": "1196.56",
  "ValorDaCorrecao": "0",
  "ValorDoJuros": "383.29",
  "ValorDoSeguro": "0",
  "ValorTaxaBancaria": "0",
  "ValorDaPrestacao": "1579.85",
  "ValorDoSaldoAnterior": "13602.6",
  "ValorDoJurosCapitalizados": "0",
  "ValorDoSaldoAtual": "12406.04"
},
{
  "NumeroDaParcela": "4",
  "DataDeVencimento": "2022-03-07T00:00:00",
  "ValorDaAmortizacao": "1230.27",
  "ValorDaCorrecao": "0",
  "ValorDoJuros": "349.58",
  "ValorDoSeguro": "0",
  "ValorTaxaBancaria": "0",
  "ValorDaPrestacao": "1579.85",
  "ValorDoSaldoAnterior": "12406.04",
  "ValorDoJurosCapitalizados": "0",
  "ValorDoSaldoAtual": "11175.77"
},
{
  "NumeroDaParcela": "5",
  "DataDeVencimento": "2022-04-07T00:00:00",
  "ValorDaAmortizacao": "1264.94",
  "ValorDaCorrecao": "0",
  "ValorDoJuros": "314.91",
```

```
"ValorDoSeguro": "0",
"ValorTaxaBancaria": "0",
"ValorDaPrestacao": "1579.85",
"ValorDoSaldoAnterior": "11175.77",
"ValorDoJurosCapitalizados": "0",
"ValorDoSaldoAtual": "9910.83"
},
{
  "NumeroDaParcela": "6",
  "DataDeVencimento": "2022-05-07T00:00:00",
  "ValorDaAmortizacao": "1300.58",
  "ValorDaCorrecao": "0",
  "ValorDoJuros": "279.27",
  "ValorDoSeguro": "0",
  "ValorTaxaBancaria": "0",
  "ValorDaPrestacao": "1579.85",
  "ValorDoSaldoAnterior": "9910.83",
  "ValorDoJurosCapitalizados": "0",
  "ValorDoSaldoAtual": "8610.25"
},
{
  "NumeroDaParcela": "7",
  "DataDeVencimento": "2022-06-07T00:00:00",
  "ValorDaAmortizacao": "1337.23",
  "ValorDaCorrecao": "0",
  "ValorDoJuros": "242.62",
  "ValorDoSeguro": "0",
  "ValorTaxaBancaria": "0",
  "ValorDaPrestacao": "1579.85",
  "ValorDoSaldoAnterior": "8610.25",
  "ValorDoJurosCapitalizados": "0",
  "ValorDoSaldoAtual": "7273.02"
},
{
  "NumeroDaParcela": "8",
  "DataDeVencimento": "2022-07-07T00:00:00",
  "ValorDaAmortizacao": "1374.91",
  "ValorDaCorrecao": "0",
  "ValorDoJuros": "204.94",
  "ValorDoSeguro": "0",
  "ValorTaxaBancaria": "0",
  "ValorDaPrestacao": "1579.85",
  "ValorDoSaldoAnterior": "7273.02",
  "ValorDoJurosCapitalizados": "0",
  "ValorDoSaldoAtual": "5898.11"
},
{
  "NumeroDaParcela": "9",
  "DataDeVencimento": "2022-08-07T00:00:00",
  "ValorDaAmortizacao": "1413.65",
```

```
"ValorDaCorrecao": "0",
"ValorDoJuros": "166.2",
"ValorDoSeguro": "0",
"ValorTaxaBancaria": "0",
"ValorDaPrestacao": "1579.85",
"ValorDoSaldoAnterior": "5898.11",
"ValorDoJurosCapitalizados": "0",
"ValorDoSaldoAtual": "4484.46"
},
{
  "NumeroDaParcela": "10",
  "DataDeVencimento": "2022-09-07T00:00:00",
  "ValorDaAmortizacao": "1453.49",
  "ValorDaCorrecao": "0",
  "ValorDoJuros": "126.36",
  "ValorDoSeguro": "0",
  "ValorTaxaBancaria": "0",
  "ValorDaPrestacao": "1579.85",
  "ValorDoSaldoAnterior": "4484.46",
  "ValorDoJurosCapitalizados": "0",
  "ValorDoSaldoAtual": "3030.97"
},
{
  "NumeroDaParcela": "11",
  "DataDeVencimento": "2022-10-07T00:00:00",
  "ValorDaAmortizacao": "1494.44",
  "ValorDaCorrecao": "0",
  "ValorDoJuros": "85.41",
  "ValorDoSeguro": "0",
  "ValorTaxaBancaria": "0",
  "ValorDaPrestacao": "1579.85",
  "ValorDoSaldoAnterior": "3030.97",
  "ValorDoJurosCapitalizados": "0",
  "ValorDoSaldoAtual": "1536.53"
},
{
  "NumeroDaParcela": "12",
  "DataDeVencimento": "2022-11-07T00:00:00",
  "ValorDaAmortizacao": "1536.53",
  "ValorDaCorrecao": "0",
  "ValorDoJuros": "43.32",
  "ValorDoSeguro": "0",
  "ValorTaxaBancaria": "0",
  "ValorDaPrestacao": "1579.85",
  "ValorDoSaldoAnterior": "1536.53",
  "ValorDoJurosCapitalizados": "0",
  "ValorDoSaldoAtual": "0"
}
]
},
```

```

    "TotalDeAmortizacao": "15898.23",
    "TotalDaTaxaDeServico": "0",
    "TotalDoSeguro": "0",
    "TotalDaTaxaDeCorrecao": "0",
    "TotalDeJuros": "3509.79",
    "TotalDoValorDasParcelas": "18958.20",
    "detalhamentoDaCET": {
      "PorcentagemDeJuros": "2.80",
      "PorcentagemDeImpostos": "0.42",
      "PorcentagemDeTarifas": "0.00",
      "PorcentagemDeServicos": "0.00"
    }
  },
  "bank": null,
  "sendVortex": "True",
  "status": "Enviada",
  "updatedAt": {
    "$date": "2022-05-03T19:28:10.454Z"
  },
  "createdAt": "2021-10-07T12:20:19.093830+00:00",
  "ccbCode": "A0988020-000",
  "emitedAt": {
    "$date": "2021-10-07T12:28:25.832Z"
  }
}

```

Dentro do documento no MongoDB, se encontram as informações do contrato e dos pagamentos deste contrato.

Desafio:

- Demonstrar qual estrutura dedados você faria para receber estes dados no Postgres.
- Escolha de ferramentas e bibliotecas que faria para resolver este problema.
- Como a solução lidaria com falhas na conexão, registros duplicados ou inconsistências?
- Defina a estratégia para execução periódica.

Desafio 2

As informações das CPRs e Notas Promissórias estão disponíveis em arquivos PDF e planilhas. você deve propor uma solução para extrair os dados destes documentos, padronizá-los e armazená-los em um ambiente de análise.

Arquivos Anexo

Desafio:

- Apresentar uma abordagem para extrair dados de PDFs.

- Propor um método para ler e tratar dados de planilhas, incluindo o armazenamento e extração.
- Unificação os dados extraídos, validando e transformando conforme necessário, e carga para uma base de dados central, destacando qual estrutura utilizaria e por que.

Desafio 3

Diariamente, um arquivo posicional(CNAB200) contendo o registro dos pagamentos é gerado. Você deverá desenvolver uma solução que faça o parsing deste arquivo, valide os dados os insira tanto no banco, quando numa solução centralizada.

Dicionário CNAB200

- **Header (1ª linha, 200 caracteres)**
 - **Posições 1–3:** 237 (código do banco)
 - **Posições 4–7:** 0001 (agência)
 - **Posições 8–17:** 2025020401 (data/hora de geração – exemplo fictício)
 - **Posições 18–37:** FINTECH EMPRESA LTDA (20 caracteres exatos)
 - **Posições 38–57:** ARQUIVO RETORNO (15 caracteres) + 5 espaços = 20
 - **Posições 58–200:** 143 espaços (filler)
- **Detalhes (linhas 2 a 9, cada uma com 200 caracteres)**
 - **Posição 1:** 1 (identifica registro de detalhe)
 - **Posições 2–4:** 237 (código do banco)
 - **Posições 5–14:** Nosso Número (10 dígitos, ex.: 0000000001)
 - **Posições 15–24:** Número do Documento (10 dígitos, ex.: 0000001001)
 - **Posições 25–32:** Data de Vencimento (20250210)
 - **Posições 33–42:** Valor do Título (0000010000 = R\$ 100,00)
 - **Posições 43–50:** Data do Pagamento (20250211)
 - **Posições 51–62:** Valor Pago (000000010000 = R\$ 100,00)
 - **Posições 63–77:** CPF/CNPJ do Pagador (15 dígitos, ex.: 0000000000000001)
 - **Posições 78–107:** Nome do Pagador (30 caracteres, alinhado à esquerda + espaços)
 - **Posições 108–200:** 93 espaços (filler)
- **Trailer (10ª linha, 200 caracteres)**
 - **Posição 1:** 9 (identifica registro trailer)
 - **Posições 2–7:** 000008 (quantidade de registros de detalhe)
 - **Posições 8–20:** 00000008000000 (soma dos valores pagos – 8 boletos de R\$ 100,00 = R\$ 800,00; em centavos = 80000 , com zeros à esquerda até 13 dígitos)
 - **Posições 21–200:** 180 espaços (filler)

```
23700012025020401FINTECH EMPRESA LTDAARQUIVO RETORNO
1237000000000100000010012025021000000100002025021100000001000000000000001JOA
```

O DA SILVA
1237000000000200000010022025021000000100002025021100000001000000000000002MAR
IA DOS SANTOS
1237000000000300000010032025021000000100002025021100000001000000000000003CAR
LOS ALMEIDA
1237000000000400000010042025021000000100002025021100000001000000000000004ANA
PEREIRA
1237000000000500000010052025021000000100002025021100000001000000000000005PED
RO SOUZA
1237000000000600000010062025021000000100002025021100000001000000000000006LUI
SA FERREIRA
1237000000000700000010072025021000000100002025021100000001000000000000007MAR
COS PEREIRA
1237000000000800000010082025021000000100002025021100000001000000000000008JUL
IA COSTA
900000800000000800000

Desafio

- **Parsing do arquivo:** Definir uma função que leia o arquivo posicional considerando a especificação.
- **Validação:** Implementar verificações.
- **Carga:** Qual estrutura faria para a carga.