



TRABALHO 02

ÁRVORES-B

Prazo para entrega: 19/05/2021 – 23:59
--

Atenção

- **Sistema de submissão:** <http://judge.sor.ufscar.br/ori/>
- **Arquivo:** deverá ser submetido um único código-fonte seguindo o padrão de nomenclatura <RA>_ORI_T02.c, ex: 123456_ORI_T02.c;
- **E/S:** tanto a entrada quanto a saída de dados devem ser de acordo com os casos de testes abertos;
- **Identificadores de variáveis:** escolha nomes apropriados;
- **Documentação:** inclua comentários e indente corretamente o programa;
- **Erros de compilação:** nota **zero** no trabalho;
- **Tentativa de fraude:** **nota zero na média para todos os envolvidos.** Fraudes, como tentativas de compras de soluções ou cópias de parte ou de todo código-fonte, de qualquer origem, implicará na reprovação direta na disciplina. Partes do código cujas **ideias** foram desenvolvidas em colaboração com outro(s) aluno(s) devem ser devidamente documentadas em comentários no referido trecho. O que **NÃO** autoriza a cópia de trechos de código, a codificação em conjunto, compra de soluções, ou compartilhamento de tela para resolução do trabalho. Portanto, compartilhem ideias em alto nível, modos de resolver o problema, mas não o código;
- **Utilize as mensagens pré-definidas (#define).**

1 Contexto

Suas apostas se provaram corretas e o MVP da **UFSPay** foi um sucesso! Milhares de usuários passaram a usar o seu produto e sua base de clientes e transações começou a crescer abruptamente.

Ao acompanhar as métricas de qualidade do seu sistema, como quantidade de requisições e tempo de resposta para cada operação, além das métricas do banco de dados, você percebeu que algumas operações começaram a ficar demasiadamente lentas. O cadastro de clientes e chaves PIX são rápidos, mas a manutenção dos índices está demorando demais; as buscas por chaves PIX demoram; as listagens também estão demandando muito tempo para retornar um resultado, entre outros problemas de lentidão. Por meio de uma análise minuciosa, foi constatado que esse gargalo deve-se ao fato de que os índices simples (listas ordenadas) não são mais adequados, pois não cabem mais na memória RAM e, portanto, demandam muitos acessos ao disco para realizar as operações.

Após fazer um estudo técnico (também conhecido como busca em fóruns da internet), você constatou que a estrutura de dados mais adequada para armazenar índices volumosos é a Árvore-B.

2 Base de dados da aplicação

Relembrando, o sistema é composto por dados de usuários e de transações conforme descrito a seguir.

2.1 Dados dos usuários

- **CPF:** identificador único de um cliente (chave primária) que contém 11 números. Não poderá existir outro valor idêntico na base de dados. Ex: 57956238064;
- **Nome completo;**
- **Data de nascimento:** data no formato <DD>/<MM>/<AAAA>. Ex: 12/01/1997;
- **E-mail;**
- **Celular:** número de contato no formato <99><999999999>. Ex: 15924835754;
- **Saldo:** saldo do cliente no formato <9999999999>.<99>. Ex: 0000004605.10;
- **Chaves PIX:** campo multivalorado separado pelo caractere '&'. O cliente poderá cadastrar até 4 chaves, apenas uma de cada tipo: C (CPF), N (número de celular), E (email) e A (aleatória). As chaves deverão ser únicas e derivadas do cadastro do usuário, exceto para a chave aleatória. No arquivo, o primeiro caractere será o tipo da chave, e as chaves estão na ordem de inserção. Ex: C57956238064&N15924835754&Ejose@gmail.com&A123e4567-e12b-12d1-a456-426655440000.

2.2 Dados das transações

- **CPF origem:** CPF do cliente que enviou o dinheiro através de uma chave PIX;

- **CPF destino:** CPF do cliente que recebeu o dinheiro através de uma chave PIX;
- **Timestamp:** data e hora em que a transação ocorreu no formato AAAAMDDHHmmSS.
Ex: 20210318143000;
- **Valor da transação:** no formato <9999999999>.<99>. Ex: 0000000042.00.

O identificador único de uma transação é uma chave composta pelo CPF de origem e *timestamp*. Garantidamente, nenhum campo de texto receberá caracteres acentuados.

2.3 Modelo Relacional e DDL

A base de dados será mantida em forma de arquivos, porém se fosse observar um modelo relacional, seria equivalente ao observado na [Figura 1](#).

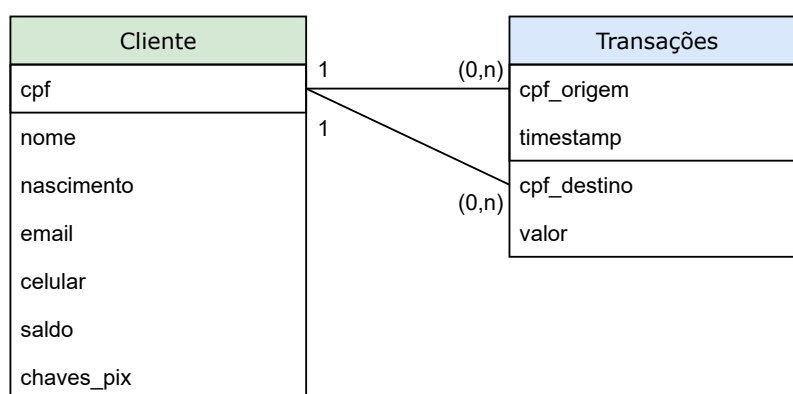


Figura 1: Modelo relacional das tabelas de transações e usuários da UFSPay

```
CREATE TABLE clientes (
  cpf          varchar(11) NOT NULL PRIMARY KEY,
  nome         text NOT NULL,
  nascimento   varchar(10) NOT NULL,
  email        text NOT NULL,
  celular      varchar(11) NOT NULL,
  saldo        numeric(12, 2) DEFAULT 0,
  chaves_pix   text[4] DEFAULT '{}')
;

CREATE TABLE transacoes (
  cpf_origem   varchar(11) NOT NULL,
  cpf_destino   varchar(11) NOT NULL,
  valor        numeric(12, 2) NOT NULL,
  timestamp    varchar(14) NOT NULL,
  PRIMARY KEY  (cpf_origem, timestamp))
;
```

Em uma implementação real, haveriam mecanismos adicionais para verificar a unicidade das chaves PIX antes da inserção de uma chave, atualização de saldo, adicionar o *timestamp* antes de inserir a transação, e situações de erro.

3 Operações suportadas pelo programa

Deve ser possível interagir com o programa através do console/terminal (modo texto) usando uma sintaxe similar à SQL, sendo que as operações a seguir devem ser fornecidas.

3.1 Cadastro

```
INSERT INTO clientes VALUES ('<CPF>', '<nome completo>', '<data de nascimento>',  
'<email>', '<celular>');
```

Uma nova conta é criada. Seu programa deve ler os seguintes campos: CPF, nome completo, data de nascimento, e-mail e celular. Inicialmente, a conta não possui saldo (R\$0,00) e não há nenhuma chave PIX cadastrada. Garantidamente, os campos serão fornecidos de maneira regular, não sendo necessário um pré-processamento da entrada. Não é permitido inserir usuários com um mesmo CPF cadastrado, portanto, caso já exista no sistema, deverá ser apresentada a mensagem padrão ERRO_PK_REPETIDA. Se a operação se concretizar, exiba a mensagem padrão SUCESSO.

Lembre-se de atualizar todos os índices necessários durante a inserção.

3.2 Depósito e saque

```
UPDATE clientes SET saldo = saldo + <valor> WHERE cpf = '<CPF do cliente>';
```

O usuário deverá ser capaz de depositar ou sacar dinheiro de uma conta informando seu CPF e o valor. Caso a conta para dado CPF não exista, seu programa deverá exibir a mensagem padrão ERRO_REGISTRO_NAO_ENCONTRADO. O valor é positivo em caso de depósito (ex: 25.40) e negativo em caso de saque (ex: -50.00). Caso o saque exceda o saldo disponível do cliente, a operação não poderá ser realizada e deverá ser apresentada a mensagem padrão ERRO_SALDO_NAO_SUFICIENTE. Caso o valor seja nulo (0.00), a operação não poderá ser realizada e deverá ser apresentada a mensagem padrão ERRO_VALOR_INVALIDO. Se a operação se concretizar, exiba a mensagem padrão SUCESSO.

3.3 Cadastro de chave PIX

```
UPDATE clientes SET chaves_pix = array_append(chaves_pix, '<tipo da chave PIX>')  
WHERE cpf = '<CPF do cliente>';
```

O usuário deverá ser capaz de inserir uma nova chave PIX para um cliente. Seu programa deverá solicitar como entrada o CPF do cliente e o tipo da chave PIX (C: CPF, N: número de celular, E: email, A: chave aleatória) e o sistema deverá gerar a chave PIX com base nos dados do cliente. O cliente poderá ter uma única chave de cada tipo. Todas as chaves devem ser únicas. Caso o cliente não exista, deverá ser apresentada a mensagem padrão ERRO_REGISTRO_NAO_ENCONTRADO. Caso o usuário tente cadastrar chaves já cadastradas por outros clientes, deverá ser apresentada a mensagem padrão ERRO_CHAVE_PIX_REPETIDA. Caso o cliente informe um tipo inválido, deve ser apresentada a mensagem padrão ERRO_TIPO_PIX_INVALIDO. Caso o cliente tente cadastrar uma chave de um tipo que ele já possui, deverá ser apresentada a mensagem padrão ERRO_CHAVE_PIX_DUPLICADA. A chave

PIX aleatória deverá ser gerada a partir da função `void new_uuid(char buffer[37])` fornecida no código base e com exemplo de uso. Se a operação se concretizar, exibir a mensagem padrão `SUCESSO`.

3.4 Transferência por chave PIX

```
INSERT INTO transacoes VALUES ('<chave PIX origem>', '<chave PIX destino>',  
<valor>);
```

O usuário deverá ser capaz de transferir dinheiro da sua conta para outra a partir das respectivas chaves PIX. Seu programa deverá solicitar como entrada a chave PIX da conta de origem, a chave PIX da conta de destino e o valor a ser transferido. Caso a chave não exista ou ambas as chaves pertençam ao mesmo CPF, deverá ser apresentada a mensagem padrão `ERRO_CHAVE_PIX_INVALIDA`. Caso a transferência exceda o saldo disponível do cliente de origem, deve ser apresentada a mensagem padrão `ERRO_SALDO_NAO_SUFICIENTE`. Caso o valor seja menor ou igual a zero, deve ser apresentada a mensagem padrão `ERRO_VALOR_INVALIDO`. Para concretizar a operação, será necessário atualizar os saldos de ambos os clientes e obter o *timestamp* do momento que ocorreu a transação utilizando a função `void current_timestamp(char buffer[15])` fornecida no código base e com exemplo de uso. Se a operação se concretizar, exibir a mensagem padrão `SUCESSO`.

3.5 Busca

O usuário deverá buscar clientes e transações pela(s) chave(s).

Antes dos resultados das buscas deverá ser impresso os RRNs dos nós percorridos. Exemplo:

Nos percorridos: 4 0 2 3

As seguintes operações de busca por clientes e transações deverão ser implementadas.

3.5.1 Clientes

O usuário deverá ser capaz de buscar clientes pelos seguintes atributos:

(a) Por CPF:

```
SELECT * FROM clientes WHERE cpf = '<CPF>';
```

Solicitar ao usuário o CPF do cliente. Caso a conta não exista, seu programa deverá exibir a mensagem padrão `ERRO_REGISTRO_NAO_ENCONTRADO`. Caso a conta exista, todos os seus dados deverão ser impressos na tela de forma formatada.

(b) Por chave PIX:

```
SELECT * FROM clientes WHERE '<chave PIX>' = ANY (chaves_pix);
```

Solicitar ao usuário uma chave PIX. Caso a chave não exista, seu programa deverá exibir a mensagem padrão `ERRO_REGISTRO_NAO_ENCONTRADO`. Caso a chave exista, todos os dados do cliente que possui a chave deverão ser impressos na tela de forma formatada.

3.5.2 Transações

O usuário deverá ser capaz de buscar transações pelos seguintes atributos:

- (a) Por CPF de origem e data específica:

```
SELECT * FROM transacoes WHERE cpf_origem = '<CPF>' AND timestamp = '<data>';
```

Solicitar ao usuário um CPF de origem e uma data no formato AAAAMDDHhmmSS. Caso a transação não exista, seu programa deverá exibir a mensagem padrão ERRO_REGISTRO_NAO_ENCONTRADO. Caso a transação exista, todos os dados da transação deverão ser impressos na tela de forma formatada.

3.6 Listagem

As seguintes operações de listagem de clientes e transações deverão ser implementadas.

3.6.1 Clientes

- (a) Por CPF:

```
SELECT * FROM clientes ORDER BY cpf ASC;
```

Exibe todos os clientes ordenados de forma crescente pelo CPF. Caso nenhum registro for retornado, seu programa deverá exibir a mensagem padrão AVISO_NENHUM_REGISTRO_ENCONTRADO.

3.6.2 Transações

- (a) Por período:

```
SELECT * FROM transacoes WHERE timestamp BETWEEN '<data início>' AND '<data fim>'
ORDER BY timestamp ASC;
```

Exibe todas as transações realizadas em um período de tempo (timestamp entre <data início> e <data fim>), em ordem cronológica. Caso nenhum registro for retornado, seu programa deverá exibir a mensagem padrão AVISO_NENHUM_REGISTRO_ENCONTRADO.

- (b) Por CPF de origem e período:

```
SELECT * FROM transacoes WHERE cpf_origem = '<CPF>'
AND timestamp BETWEEN '<data início>' AND '<data fim>'
ORDER BY timestamp ASC;
```

Exibe todas as transações realizadas por um CPF (cpf_origem) em um período de tempo (timestamp entre <data início> e <data fim>), em ordem cronológica. Caso nenhum registro for retornado, seu programa deverá exibir a mensagem padrão AVISO_NENHUM_REGISTRO_ENCONTRADO.

3.7 Imprimir arquivos de dados

O sistema deverá imprimir os arquivos de dados da seguinte maneira:

- (a) Dados dos clientes:

```
\echo file ARQUIVO_CLIENES
```

Imprime o arquivo de dados de clientes. Caso esteja vazio, apresentar a mensagem padrão ERRO_ARQUIVO_VAZIO;

- (b) Dados das transações:

```
\echo file ARQUIVO_TRANSACOES
```

Imprime o arquivo de dados de transações. Caso esteja vazio, apresentar a mensagem padrão ERRO_ARQUIVO_VAZIO.

3.8 Imprimir arquivos de índice

O sistema deverá imprimir os arquivos de índices da seguinte maneira:

- (a) Índice de clientes:

```
\echo index clientes_idx
```

Imprime o arquivo de índice primário para clientes (`clientes_idx`). Caso o arquivo esteja vazio, imprimir ERRO_ARQUIVO_VAZIO;

- (b) Índice de transações:

```
\echo index transacoes_idx
```

Imprime o arquivo de índice primário para transações (`transacoes_idx`). Caso o arquivo esteja vazio, imprimir ERRO_ARQUIVO_VAZIO;

- (c) Índice de chaves PIX:

```
\echo index chaves_pix_idx
```

Imprime o arquivo de índice secundário para chaves PIX (`chaves_pix_idx`). Caso o arquivo esteja vazio, imprimir ERRO_ARQUIVO_VAZIO;

- (d) Índice de transações:

```
\echo index timestamp_cpf_origem_idx
```

Imprime o arquivo de índice secundário para o *timestamp* e CPF do cliente de origem das transações (`timestamp_cpf_origem_idx`). Caso o arquivo esteja vazio, imprimir ERRO_ARQUIVO_VAZIO.

3.9 Finalizar

\q

Libera memória eventualmente alocada e encerra a execução do programa.

4 Arquivos de dados

Novamente, nenhum arquivo ficará salvo em disco. O arquivo de dados e os de índices serão simulados em *strings* e os índices serão sempre criados na inicialização do programa e manipulados em memória RAM até o término da execução. Suponha que há espaço suficiente em memória RAM para todas as operações.

Deve-se utilizar as variáveis globais `ARQUIVO_CLIENTES` e `ARQUIVO_TRANSACOES` e as funções de leitura e escrita em *strings*, como `sprintf` e `sscanf`, para simular as operações de leitura e escrita em arquivo. Os arquivos de dados devem ser no formato ASCII (arquivo texto).

`ARQUIVO_CLIENTES`: deverá ser organizado em registros de tamanho fixo de 256 *bytes* (256 caracteres). Os campos nome, e-mail e chaves PIX devem ser de tamanho variável. Os demais campos devem ser de tamanho fixo: CPF (11 *bytes*), data de nascimento (10 *bytes*), celular (11 *bytes*) e saldo (13 *bytes*). Portanto, os campos de tamanho fixo de um registro ocuparão 45 *bytes*. Os campos do registro devem ser separados pelo caractere delimitador ‘;’ (ponto e vírgula), cada registro terá 7 delimitadores (um para cada campo). O campo multivalorado será separado pelo caractere ‘&’. Caso o registro tenha menos de 256 *bytes*, o espaço remanescente deverá ser preenchido com o caractere ‘#’ de forma a completar os 256 *bytes*. Como são 45 *bytes* fixos + 7 *bytes* de delimitadores, então os campos variáveis devem ocupar no máximo 204 *bytes* (incluindo os delimitadores do campo multivalorado), para evitar que o registro exceda 256 *bytes*.

Exemplo de arquivo de dados de clientes:

```
44535687915;JOSE AUGUSTO;12/11/1951;joseaugusto@hotmail.com;1599
6587458;0000042580.80;C44535687915&Ejoseaugusto@hotmail.com;####
#####
#####
14578965815;MARIANA DIAS;18/05/1995;marianadias@gmail.com;159817
54878;0000021850.00;N15981754878&C14578965815;#####
#####
#####
34672961815;CARLA AMARAL;20/03/1990;carlaamaral@gmail.com;159917
34220;0000311050.20;;#####
#####
#####
```

`ARQUIVO_TRANSACOES`: o arquivo de transações deverá ser organizado em registros de tamanho fixo de 49 *bytes* (49 caracteres). Todos os campos possuem tamanho fixo: CPF de origem (11 *bytes*), CPF de destino (11 *bytes*), valor (13 *bytes*) e *timestamp* (14 *bytes*).

Exemplo de arquivo de dados das transações:

```
44535687915 14578965815 0000000025.40 20150203142345
44535687915 14578965815 0000000142.60 20150405123522
14578965815 44535687915 0000001045.90 20150615114802
```

Como os campos de transações possuem tamanho fixo, não são necessários delimitadores (foram inseridos espaços para maior clareza).

Note que não há quebras de linhas nos arquivos (elas foram inseridas aqui apenas para facilitar a visualização da sequência de registros).

5 Instruções para as operações com os registros

- **Inserção:** cada cliente e transação devem ser inseridos no final de seus respectivos arquivos de dados, e atualizados os índices.
- **Atualização:** o único campo alterável é o saldo. O registro deverá ser localizado acessando o índice primário e o novo saldo deverá ser atualizado no registro na mesma posição em que está (não deve ser feita remoção seguida de inserção). Note que o campo de saldo sempre ocupará 13 *bytes*.

6 Criação de índices

No cenário atual, há um grande volume de dados, e nem mesmo os índices cabem em RAM, portanto, para simular essa situação, a única informação que você deverá manter todo tempo em memória, é o RRN da raiz de cada Árvore-B. Assuma que um nó do índice corresponde a uma página e, portanto, cabe no *buffer* de memória. Dessa forma, trabalhe apenas com a menor quantidade de nós necessários das árvores por vez, pois isso implica em reduzir a quantidade de *seeks* e de informação transferida entre as memórias primária e secundária. **É terminantemente proibido manter uma cópia dos índices inteiros em Tipos Abstratos de Dados (TADs) que não Árvores-B.**

Todas as árvores terão a mesma ordem (m) e **a promoção deverá ser sempre pelo sucessor imediato** (menor chave da sub-árvore à direita). Todo novo nó criado deverá ser inserido no final do respectivo arquivo de índice.

6.1 Índices primários

6.1.1 `clientes_idx`

Índice primário do tipo Árvore-B que contém o CPF do cliente (chave primária) e o RRN do respectivo registro no arquivo de dados, ordenado pela chave primária.

Cada registro da Árvore-B para o índice `clientes_idx` é composto por:

- *Contador de chaves:* 3 bytes para a quantidade de chaves;

- *Chaves ordenadas:* $(m - 1) * (11 \text{ bytes de chave primária} + 4 \text{ bytes para o RRN do arquivo de dados})$ bytes para armazenar as chaves. Para as chaves não usadas, preencha todos os bytes com '#';
- *Indicador de folha:* 1 byte para indicar se o nó é folha 'T' (True) ou não 'F' (False);
- *Apontadores para os filhos:* $(m * 3 \text{ bytes para cada RRN filho})$ bytes para indicar os RRNs dos nós descendentes do nó atual. Note que esse RRN se refere ao próprio arquivo de índice primário. Utilize '***' para indicar que aquela posição do vetor de descendentes é nula.

Exemplo da representação da Árvore-B de ordem 3, após a inserção das chaves: 43487689087, 52587909876 e 21046578965 (nesta ordem).

- Inserindo chave 43487689087:

⁰ 43487689087

Em disco, o arquivo de índice primário seria:

001 | 43487689087 0000 | ##### ##### | T | *** *** ***

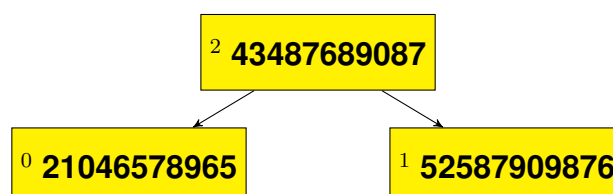
- Inserindo chave 52587909876:

⁰ 43487689087, 52587909876

Em disco:

002 | 43487689087 0000 | 52587909876 0001 | T | *** *** ***

- Inserindo chave 21046578965:



Ao inserir a chave iniciada por 21046578965, ocorre um *overflow* na raiz, sendo necessário criar então, 2 novos nós (de RRN 1 e 2 respectivamente). O primeiro, será para a redistribuição de chaves, e o segundo para receber a promoção de chave que será a nova raiz. Portanto:

001 | 21046578965 0002 | ##### ##### | T | *** *** ***
 001 | 52587909876 0001 | ##### ##### | T | *** *** ***
 001 | 43487689087 0000 | ##### ##### | F | 000 001 ***

Em resumo, um nó da Árvore-B de CPF, de ordem m , possui as seguintes informações:

```
{QUANTIDADE DE CHAVES}
{CPF_1} {RRN_1}
{CPF_2} {RRN_2}
...
{CPF_(m-1)} {RRN_(m-1)}
{FOLHA}
{RRN FILHO_1} {RRN FILHO_2} ... {RRN FILHO_m}
```

Note que, aqui também não há quebras de linhas no arquivo, espaços ou *pipes* ('|'), Eles foram inseridas apenas para exemplificar a sequência de registros.

6.1.2 transacoes_idx

Índice primário do tipo Árvore-B que contém o CPF de origem da transação (em ordem crescente) e o *timestamp* (em ordem cronológica), e o RRN respectivo do registro no arquivo de transações.

Cada registro da Árvore-B para o índice `transacoes_idx` é composto por:

- *Contador de chaves*: 3 bytes para a quantidade de chaves;
- *Chaves ordenadas*: $(m - 1) * (25 \text{ bytes de chave primária [11 do CPF + 14 do timestamp]} + 4 \text{ bytes para o RRN do arquivo de dados})$ bytes para armazenar as chaves. Para as chaves não usadas, preencha todos os bytes com '#';
- *Indicador de folha*: 1 byte para indicar se o nó é folha 'T' (True) ou não 'F' (False);
- *Apontadores para os filhos*: $(m * 3 \text{ bytes para cada RRN filho})$ bytes para indicar os RRNs dos nós descendentes do nó atual. Note que esse RRN se refere ao próprio arquivo de índice primário. Utilize '***' para indicar que aquela posição do vetor de descendentes é nula.

Exemplo de arquivo de índice primário de transações representado por uma Árvore-B de ordem 4, após a inserção das chaves na ordem: '44678965437 20210325091500 0002', '34678965321 20210912151010 0000' e '51478965098 20210101183012 0001'.

⁰ 346..., 446..., 514...

```
003 | 34678965321 20210912151010 0002 | 44678965437 20210325091500 0000 |
    51478965098 20210101183012 0001 | T | *** *** *** ***
```

Em resumo, um nó da Árvore-B de transações, de ordem m , possui as seguintes informações:

```
{QUANTIDADE DE CHAVES}
{CPF_1} {TIMESTAMP_1} {RRN_1}
{CPF_2} {TIMESTAMP_2} {RRN_2}
...
{CPF_(m-1)} {TIMESTAMP_(m-1)} {RRN_(m-1)}
{FOLHA}
{RRN FILHO_1} {RRN FILHO_2} ... {RRN FILHO_m}
```

Novamente, não há quebras de linhas no arquivo. Elas foram inseridas apenas para exemplificar a sequência de registros.

6.2 Índices secundários

6.2.1 `chaves_pix_idx`

Este índice secundário do tipo Árvore-B contém as chaves PIX de todos os clientes (em ordem lexicográfica) e a chave primária do respectivo cliente.

Cada registro da Árvore-B para o índice `chaves_pix_idx` é composto por:

- *Contador de chaves*: 3 bytes para a quantidade de chaves;
- *Chaves ordenadas*: $(m - 1) * (47 \text{ bytes de chave PIX} + 11 \text{ bytes de chave primária})$ bytes para armazenar as chaves. Para as chaves não usadas ou para completar o tamanho da chave PIX, preencha todos os bytes com '#';
- *Indicador de folha*: 1 byte para indicar se o nó é folha 'T' (True) ou não 'F' (False);
- *Apontadores para os filhos*: $(m * 3 \text{ bytes para cada RRN filho})$ bytes para indicar os RRNs dos nós descendentes do nó atual. Note que assim como no índice primário, esse RRN se refere ao próprio arquivo de índice secundário. Utilize '***' para indicar que aquela posição do vetor de descendentes é nula.

Exemplo de arquivo de índice secundário de chaves PIX, representado por uma Árvore-B de ordem 5, após a inserção da chave 'maria@gmail.com 21046578965'.

0 maria@gmail.com...

```
001 | maria@gmail.com##### 21046578965 |
##### |
##### |
##### |
T | *** *** *** *** ***
```

Em resumo, um nó da Árvore-B de chaves PIX, de ordem m , possui as seguintes informações:

```
{QUANTIDADE DE CHAVES}
{CHAVE_PIX_1} {CPF_1}
{CHAVE_PIX_2} {CPF_2}
...
{CHAVE_PIX_(m-1)} {CPF_(m-1)}
{FOLHA}
{RRN FILHO_1} {RRN FILHO_2} ... {RRN FILHO_m}
```

Lembre-se, aqui também não há quebras de linhas no arquivo, espaços ou *pipes* ('|'), Eles foram inseridas apenas para exemplificar a sequência de registros.

6.2.2 timestamp_cpf_origem_idx

Índice secundário do tipo Árvore-B que contém o *timestamp* das transações (em ordem cronológica) e o CPF do cliente de origem (em ordem crescente). Note que esses campos já compõem a chave primária da transação.

Cada registro da Árvore-B para o índice `timestamp_cpf_origem_idx` é composto por:

- *Contador de chaves*: 3 bytes para a quantidade de chaves;
- *Chaves ordenadas*: $(m - 1) * (25 \text{ bytes de chave primária } [14 \text{ do } timestamp + 11 \text{ do CPF}])$ bytes para armazenar as chaves. Para as chaves não usadas, preencha todos os bytes com '#';
- *Indicador de folha*: 1 byte para indicar se o nó é folha 'T' (True) ou não 'F' (False);
- *Apontadores para os filhos*: $(m * 3 \text{ bytes para cada RRN filho})$ bytes para indicar os RRNs dos nós descendentes do nó atual. Note que assim como no índice primário, esse RRN se refere ao próprio arquivo de índice secundário. Utilize '***' para indicar que aquela posição do vetor de descendentes é nula.

Exemplo de arquivo de índice primário representado por uma Árvore-B de ordem 4, após a inserção das chaves na ordem: '20210325091500 44678965437 0002', '20210912151010 34678965321 0000' e '20210101183012 51478965098 0001'.

⁰ **20210101183012..., 20210325091500..., 20210912151010...**

003 | 20210101183012 51478965098 0001 | 20210325091500 44678965437 0000 |
20210912151010 34678965321 0002 | T | *** *** *** ***

Em resumo, um nó da Árvore-B de transações, de ordem m , possui as seguintes informações:

```
{QUANTIDADE DE CHAVES}  
{TIMESTAMP_1} {CPF_1}  
{TIMESTAMP_2} {CPF_2}  
...  
{TIMESTAMP_(m-1)} {CPF_(m-1)}  
{FOLHA}  
{RRN FILHO_1} {RRN FILHO_2} ... {RRN_FILHO_m}
```

Novamente, não há quebras de linhas no arquivo, espaços ou *pipes* ('|'), Eles foram inseridos apenas para exemplificar a sequência de registros.

Deverá ser desenvolvida uma rotina para a criação de cada índice. Os índices serão sempre criados e manipulados utilizando os pseudo-arquivos de índices na inicialização e liberados ao término do programa. Note que o ideal é que os índices primários sejam criados primeiro, depois os secundários.

7 Inicialização do programa

Para que o programa inicie corretamente, deve-se realizar o seguinte procedimento:

1. Inserir o comando `SET BTREE_ORDER '<ORDEM DAS ÁRVORES-B>'`; para informar a ordem das Árvores-B. Caso não informado, é definido como 3 por padrão;
2. Inserir o comando `SET ARQUIVO_CLIENTES '<DADOS DE CLIENTES>'`; para informar os dados contidos no arquivo de clientes ou `SET ARQUIVO_CLIENTES ''`; caso o arquivo esteja vazio;
3. Inserir o comando `SET ARQUIVO_TRANSACOES '<DADOS DE TRANSAÇÕES>'`; para informar os dados contidos no arquivo de transações ou `SET ARQUIVO_TRANSACOES ''`; caso o arquivo esteja vazio;
4. Inicializar as estruturas de dados dos índices.

8 Implementação

Implemente suas funções utilizando o código-base fornecido. **Não é permitido modificar os trechos de código pronto ou as estruturas já definidas.** Ao imprimir um registro, utilize as funções `exibir_cliente(int rrn)` ou `exibir_transacao(int rrn)`.

Note que são quatro índices do tipo Árvore-B. **Não é necessário implementar rotinas específicas para a manipulação de cada árvore.** Lembre-se de que funções como `bsearch` e `qsort`, por exemplo, são totalmente genéricas e funcionam com qualquer vetor, basta que seja informado o tamanho de cada elemento, o número de posições do vetor e a função de comparação. Use esse mesmo conceito nas funções de manutenção de Árvores-B, lembrando que a única informação que se altera entre elas é o tamanho das chaves.

Implementar rotinas que contenham obrigatoriamente as seguintes funcionalidades:

- Estruturas de dados adequadas para armazenar os índices em arquivos simulados por meio de *strings*;
- Verificar se os arquivos de dados existem;
- Criar os índices primários: deve refazer os índices primários a partir dos arquivos de dados;
- Criar os índices secundários: deve refazer os índices secundários a partir dos arquivos de dados;
- Inserir um registro: modifica os arquivos de dados e de índices;
- Buscar por registro: busca um registro por sua(s) chave(s) primária(a);
- Alterar um registro: modifica o campo do registro diretamente no arquivo de dados;
- Listar registros: listar os registros ordenados pelo intervalo de chaves dadas.

9 Resumo de alterações em relação ao T01

Para facilitar o desenvolvimento do T02, é possível reutilizar as funções já implementadas no T01, com ênfase nas seguintes alterações:

1. As operações de remoção de clientes e liberar espaço foram removidas;
2. Todos os índices serão do tipo Árvore-B;
3. As buscas devem conter, no começo da saída, todos os nós percorridos;
4. A ordem das Árvores-B é informada na inicialização do programa e é única para todas as árvores;
5. Para facilitar a leitura da saída do programa, o comando SQL executado também será impresso antes da saída.

10 Dicas

- Ao ler uma entrada, tome cuidado com caracteres de quebra de linha (`\n`) não capturados;
- Você nunca deve perder a referência do começo do arquivo, então não é recomendável percorrer a *string* diretamente pelo ponteiro `ARQUIVO`. Um comando equivalente a `fseek(f, 256, SEEK_SET)` é `char *p = ARQUIVO + 256;`
- Diferentemente do `fscanf`, o `sscanf` não movimenta automaticamente o ponteiro após a leitura;
- O `sprintf` adiciona automaticamente o caractere `\0` no final da *string* escrita. Em alguns casos você precisará sobrescrever a posição manualmente. Você também pode utilizar o comando `strncpy` para escrever em *strings*, esse comando, diferentemente do `sprintf`, não adiciona o caractere nulo no final;
- A função `strtok` permite navegar nas *substrings* de uma certa *string* dado o(s) delimitador(es). Porém, tenha em mente que ela deve ser usada em uma cópia da *string* original, pois ela modifica o primeiro argumento.

Before there were computers, there were algorithms.
(*Antes de haver computadores, haviam algoritmos*)
— *Introduction to Algorithms, CLRS*