

Teste 3 - Bancos de Dados

Tarefas de Preparação (podem ser feitas manualmente):

Baixar os arquivos dos últimos 2 anos no repositório público:

http://ftp.dadosabertos.ans.gov.br/FTP/PDA/demonstracoes_contabeis/

Baixar csv anexo:

Relatorio_cadop(1) (esta em anexo no e-mail)

As tarefas abaixo devem ser executadas em código na linguagem SQL em um banco PostgreSQL ou MySQL

Criar queries para criar tabelas com as colunas necessárias para o arquivo csv.

Queries de load: criar as queries para carregar o conteúdo dos arquivos obtidos nas tarefas de preparação

Atenção ao encoding dos arquivos no momento da importação!

Montar uma query analítica que traga a resposta para as seguintes perguntas:

Quais as 10 operadoras que mais tiveram despesas com "EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSISTÊNCIA A SAÚDE MEDICO HOSPITALAR" no último trimestre?

Quais as 10 operadoras que mais tiveram despesas com "EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSISTÊNCIA A SAÚDE MEDICO HOSPITALAR" no último ano?

O primeiro passo foi baixar e analisar os documentos.csv disponibilizados no site:

http://ftp.dadosabertos.ans.gov.br/FTP/PDA/demonstracoes_contabeis/

Assim que tive acesso aos documentos.csv, analisei como os dados estavam estruturados (colunas, delimitadores, tipos de dados, etc.). Verifiquei que alguns documentos estavam com "encodings" diferentes entre si; assim, tomei a liberdade de converter TODOS os documentos (relatorio_cadop.csv e os arquivos dos últimos 2 anos no repositório público) para o formato "utf-8" utilizando o editor de textos "Notepad++".

Após convertidos, alguns documentos apresentaram o delimitador ";", e outros, além do delimitador, estavam contidos em aspas duplas, recurso comum entre os arquivos.csv.

Foi possível observar que, no Excel, os documentos apresentavam símbolos desconhecidos no lugar de letras com acentuação, "ç" e outros, em virtude da não-conversão do "encoding" para utf-8.

Utilizando o XAMPP, instalei o Apache, o MySQL, configurei o acesso, adicionei as variáveis de ambiente para operar também pelo Windows PowerShell e instalei também o PhpMyAdmin.

- Criei o banco de dados "cadop", seguindo o padrão pelo nome do documento (relatorio_cadop.csv).
- Criei 2 tabelas: relatorio, dados.
- A tabela "relatorio" foi responsável por armazenar os dados do documento "relatorio_cadop.csv".

A tabela "dados" foi responsável por armazenar os dados de todos os documentos pertinentes aos últimos 2 anos do repositório público.

Os arquivos com os dados dos últimos 2 anos identificados são:

1T2020.csv

1T2021.csv

2T2020.csv

2T2021.csv

3T2020.csv

3T2021.csv

4T2020.csv

Estes documentos também passaram pela conversão de "encoding" para "utf-8".

O documento "dicionario_demonstracoes_contabeis.ods", encontrado no repositório público, apresenta algumas observações com relação aos tipos de dados e tamanho dos campos.

Me orientei em partes pelo conteúdo deste documento e tomei as decisões de tamanhos de campos e tipos de dados do banco de dados estipulando tamanhos de campo um pouco maiores.

Os tipos de dados dos campos do banco de dados foram levados em conta seguindo o documento "dicionario_demonstracoes_contabeis.ods".

Pude verificar que os dados dos últimos 2 anos tinham uma relação muito próxima com as especificações do documento "dicionario_demonstracoes_contabeis.ods".

Assim, tomei a liberdade de mesclar tais informações com a identificação mediante os dados apresentados nos documentos.csv.

É possível identificar que existe um dado em comum nas tabelas (o "registro_ans" na tabela "relatorio" e "reg_ans" na tabela "dados").

- Decidi que a chave primária da tabela "relatorio" seria "registroans" (int (8)).
- A chave estrangeira na tabela "dados" é, portanto, "reg_ans" (int (8)).
- Assim foi possível relacionar as 2 tabelas.

O próximo passo foi criar as Queries para as solicitações do teste.

Para criar as queries tomei a liberdade de interpretar a solicitação mediante os dados que foram disponibilizados nos arquivos csv.

Queries para construir a tabela "dados":

```
CREATE TABLE `dados` (  
  `data` date NOT NULL,  
  `reg_ans` int(8) NOT NULL,  
  `cd_conta_contabil` int(10) NOT NULL,  
  `descricao` text NOT NULL,  
  `vl_saldo_inicial` double NOT NULL,  
  `vl_saldo_final` double NOT NULL  
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```

Queries para a construção da tabela "relatorios":

```
CREATE TABLE `relatorio` (  
  `registroans` int(8) NOT NULL,  
  `cnpj` varchar(20) NOT NULL,  
  `razaosocial` text NOT NULL,  
  `nomefantasia` text NOT NULL,  
  `modalidade` text NOT NULL,  
  `logradouro` text NOT NULL,  
  `numero` text NOT NULL,  
  `complemento` text NOT NULL,  
  `bairro` text NOT NULL,  
  `cidade` text NOT NULL,  
  `uf` varchar(2) NOT NULL,  
  `cep` varchar(10) NOT NULL,  
  `ddd` int(3) NOT NULL,  
  `telefone` varchar(10) NOT NULL,  
  `fax` varchar(10) NOT NULL,  
  `enderecoeletronico` text NOT NULL,  
  `representante` text NOT NULL,  
  `cargorepresentante` text NOT NULL,  
  `dataregistroans` date NOT NULL  
) ENGINE=InnoDB DEFAULT CHARSET=utf8;  
  
ALTER TABLE `dados`  
  ADD KEY `extrangeira` (`reg_ans`);  
  
ALTER TABLE `relatorio`  
  ADD PRIMARY KEY (`registroans`);  
  
ALTER TABLE `dados`  
  ADD CONSTRAINT `extrangeira` FOREIGN KEY (`reg_ans`) REFERENCES `relatorio`  
  (`registroans`) ON DELETE NO ACTION ON UPDATE NO ACTION;  
  
COMMIT;
```

Para fins de organização e conferência dos resultados obtidos, estruturei uma consulta SQL para exibir:

```
Razao_Social  
Registro_ANS  
Quantidade de ocorrências da descrição solicitada por registro_ans  
Data  
Descrição  
Soma de (Saldo Inicial) e (Saldo_Final) dos resultados obtidos pelo agrupamento de  
reg_ans,  
Despesas (saldo_final - saldo_inicial)  
E outra query, com menos detalhes, para exibir:  
R.razaosocial as Razao_Social  
D.data  
D.descricao as Descricao  
Despesas
```

Admitindo que as despesas sejam resultado da operação "saldo_final - saldo_inicial", decidi relacionar as 2 tabelas com JOIN, pesquisando pela descrição "EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSISTÊNCIA A SAÚDE MEDICO HOSPITALAR", referentes ao último trimestre do ano em que houve registros (visto que os últimos registros são de 2021-10-01, considereei que a busca poderia ser entre as datas de 2021-12-31 e 2021-08-01), com saldo_final maior que 0 e exibindo os resultados do valor maior para o menor, limitando os primeiros 10 resultados.

Solicitação 1: Quais as 10 operadoras que mais tiveram despesas com "EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSISTÊNCIA A SAÚDE MEDICO HOSPITALAR" no último trimestre?

Query com um nível de detalhamento razoável:

```
SELECT
R.razaosocial as Razao_Social,
D.reg_ans as Registro_ANS_Dados,
count(D.reg_ans) as qt_regitros_na_tabela_dados,
D.data,
D.descricao as Descricao,
SUM(D.vl_saldo_inicial) as Saldo_Inicial,
SUM(D.vl_saldo_final) as Saldo_Final,
(SUM(D.vl_saldo_final) - SUM(D.vl_saldo_inicial)) as Despesas
FROM relatorio R
JOIN dados as D ON R.registroans = D.reg_ans
WHERE
D.descricao = "EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSISTÊNCIA A SAÚDE MEDICO HOSPITALAR"
AND
D.data BETWEEN '2021-08-01' AND '2021-12-31'
AND
D.vl_saldo_final > 0
GROUP BY D.reg_ans
ORDER BY Despesas DESC
LIMIT 10
```

Resultado:

Razao_Social	Registro_ANS_Dados	qt_regitros_na_tabela_dados	data	Descricao	Saldo_Inicial	Saldo_Final	Despesas
BRADESCO SAÚDE S.A.	5711	5	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	18201210231	24582805244	6381595013
AMIL ASSISTÊNCIA MÉDICA INTERNACIONAL S.A.	326305	6	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	12524650314	16929337656	4404687342
SUL AMERICA COMPANHIA DE SEGURO SAÚDE	6246	5	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	12153290986	16460778594	4337487608
NOTRE DAME INTERMÉDICA SAÚDE S.A.	359017	4	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	6002345422	7925283286	1922937884
HAPVIDA ASSISTENCIA MEDICA S.A.	368253	7	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	3751598591	5424824647	1673228056
CAIXA DE ASSISTÊNCIA DOS FUNCIONÁRIOS DO BANCO DO ...	346659	4	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	4037303284	5576445703	1539142419
CENTRAL NACIONAL UNIMED - COOPERATIVA CENTRAL	339679	6	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	3690067924	4950398326	1260330402
PREVENT SENIOR PRIVATE OPERADORA DE SAÚDE LTDA	302147	4	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	3225333142	4410724605	1185391663
UNIMED BELO HORIZONTE COOPERATIVA DE TRABALHO MÉDI...	343889	6	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	2592130226	3688610409	1096480183
UNIMED-RIO COOPERATIVA DE TRABALHO MEDICO DO RIO D...	393321	4	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	3168006357	4185369046	1017360689

Query com menos detalhes:

```
SELECT
R.razaosocial as Razao_Social,
D.data,
D.descricao as Descricao,
(SUM(D.vl_saldo_final) - SUM(D.vl_saldo_inicial)) as Despesas
FROM relatorio R
JOIN dados as D ON R.registroans = D.reg_ans
WHERE
D.descricao = "EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSISTÊNCIA A SAÚDE MEDICO
HOSPITALAR"
AND
D.data BETWEEN '2021-08-01' AND '2021-12-31'
AND
D.vl_saldo_final > 0
GROUP BY D.reg_ans
ORDER BY Despesas DESC
LIMIT 10
```

Resultado:

Razao_Social	data	Descricao	Despesas 1
BRADESCO SAÚDE S.A.	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	6381595013
AMIL ASSISTÊNCIA MÉDICA INTERNACIONAL S.A.	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	4404687342
SUL AMERICA COMPANHIA DE SEGURO SAÚDE	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	4337487608
NOTRE DAME INTERMÉDICA SAÚDE S.A.	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	1922937884
HAPVIDA ASSISTENCIA MEDICA S.A.	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	1673228056
CAIXA DE ASSISTÊNCIA DOS FUNCIONÁRIOS DO BANCO DO ...	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	1539142419
CENTRAL NACIONAL UNIMED - COOPERATIVA CENTRAL	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	1260330402
PREVENT SENIOR PRIVATE OPERADORA DE SAÚDE LTDA	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	1185391663
UNIMED BELO HORIZONTE COOPERATIVA DE TRABALHO MÉDI...	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	1096480183
UNIMED-RIO COOPERATIVA DE TRABALHO MEDICO DO RIO D...	2021-10-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	1017360689

Solicitação 2: Quais as 10 operadoras que mais tiveram despesas com "EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSISTÊNCIA A SAÚDE MEDICO HOSPITALAR" no último ano?

Query com um nível de detalhamento razoável:

```
SELECT
R.razaosocial as Razao_Social,
D.reg_ans as Registro_ANS_Dados,
count(D.reg_ans) as qt_regitros_na_tabela_dados,
D.data,
D.descricao as Descricao,
SUM(D.vl_saldo_inicial) as Saldo_Inicial,
SUM(D.vl_saldo_final) as Saldo_Final,
(SUM(D.vl_saldo_final) - SUM(D.vl_saldo_inicial)) as Despesas
FROM relatorio R
JOIN dados as D ON R.registroans = D.reg_ans
WHERE
D.descricao = "EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSISTÊNCIA A SAÚDE MEDICO
HOSPITALAR"
AND
D.data BETWEEN '2021-01-01' AND '2021-12-31'
AND
D.vl_saldo_final > 0
GROUP BY D.reg_ans
ORDER BY Despesas DESC
LIMIT 10
```

Resultado:

Razao_Social	Registro_ANS_Dados	qt_regitros_na_tabela_dados	data	Descricao	Saldo_Inicial	Saldo_Final	Despesas 1
BRADESCO SAÚDE S.A.	5711	20	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	18201210231	60428696471	42225488240
AMIL ASSISTÊNCIA MÉDICA INTERNACIONAL S.A.	326305	24	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	12524650314	41629532289	29104881975
SUL AMERICA COMPANHIA DE SEGURO SAÚDE	6246	20	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	12153290988	40045086565	27891795579
NOTRE DAME INTERMÉDICA SAÚDE S.A.	359017	16	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	6002345422	19865279832	13862934410
CAIXA DE ASSISTÊNCIA DOS FUNCIONÁRIOS DO BANCO DO ...	346659	16	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	4037303284	13253439487	9216133203
HAPVIDA ASSISTENCIA MEDICA S.A.	368253	25	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	3751568591	12771762155	9020163964
CENTRAL NACIONAL UNIMED - COOPERATIVA CENTRAL	339679	16	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	3690067924	12292938182	8602870258
PREVENT SENIOR PRIVATE OPERADORA DE SAÚDE LTDA	302147	16	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	3225333142	10728373626	7503040484
UNIMED BELO HORIZONTE COOPERATIVA DE TRABALHO MÉDI...	343889	24	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	2592130226	8720857535	6128727309
GEAP AUTOGESTÃO EM SAÚDE	323080	16	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	2113073747	7439093816	5326020069

Query com menos detalhes:

```
SELECT
R.razaosocial as Razao_Social,
D.data,
D.descricao as Descricao,
(SUM(D.vl_saldo_final) - SUM(D.vl_saldo_inicial)) as Despesas
FROM relatorio R
JOIN dados as D ON R.registroans = D.reg_ans
WHERE
D.descricao = "EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSISTÊNCIA A SAÚDE MEDICO
HOSPITALAR"
AND
D.data BETWEEN '2021-01-01' AND '2021-12-31'
AND
D.vl_saldo_final > 0
GROUP BY D.reg_ans
ORDER BY Despesas DESC
LIMIT 10
```

Resultado:

Razao_Social	data	Descricao	Despesas ▾ 1
BRADESCO SAÚDE S.A.	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	42225486240
AMIL ASSISTÊNCIA MÉDICA INTERNACIONAL S.A.	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	29104881975
SUL AMERICA COMPANHIA DE SEGURO SAÚDE	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	27891795579
NOTRE DAME INTERMÉDICA SAÚDE S.A.	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	13862934410
CAIXA DE ASSISTÊNCIA DOS FUNCIONÁRIOS DO BANCO DO ...	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	9216133203
HAPVIDA ASSISTENCIA MEDICA S.A.	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	9020163564
CENTRAL NACIONAL UNIMED - COOPERATIVA CENTRAL	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	8602870258
PREVENT SENIOR PRIVATE OPERADORA DE SAÚDE LTDA	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	7503040484
UNIMED BELO HORIZONTE COOPERATIVA DE TRABALHO MÉDI...	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	6128727309
GEAP AUTOGESTÃO EM SAÚDE	2021-01-01	EVENTOS/ SINISTROS CONHECIDOS OU AVISADOS DE ASSI...	5326020069

As queries são extensas e demandam bastante processamento, porém, ao meu ver, os resultados foram obtidos conforme solicitado.

Verifiquei quantas linhas existem nos documentos.csv referentes aos últimos 2 anos:

```
1T2020.csv - 679.641 linhas
1T2021.csv - 688.888 linhas
2T2020.csv - 672.240 linhas
2T2021.csv - 723.537 linhas
3T2020.csv - 686.757 linhas
3T2021.csv - 742.530 linhas
4T2020.csv - 865.172 linhas
4T2021.csv - 907.100 linhas
```

Ao todo foram identificadas 5.965.865 linhas nos documentos.csv. Subtraindo este número pelo número de registros do banco de dados (5.965.865 - 5.965.857) obtém-se 8, que corresponde à 1ª linha de cada documento.csv, que foram ignoradas na importação por se tratar do índice de cada coluna.

Identifiquei uma **possível** (e pequena) inconsistência nos dados dos últimos 2 anos com relação à tabela “relatorio”. Assim que criei as tabelas, entendendo que existia uma relação entre os dados “registroans” da tabela “relatorio” e “reg_ans” da tabela “dados”, fiz uma consulta para verificar se existiam registros na tabela “dados” que não estivessem relacionados com a tabela “relatorios”:

```
SELECT * FROM `dados` WHERE `reg_ans` NOT IN (SELECT relatorio.registroans FROM relatorio) GROUP BY `reg_ans`
```

Foram identificados 70 registros na tabela “dados” que não existem na tabela “relatorio”.

✓ A mostrar registros de 0 - 69 (70 total, A consulta demorou 86,8427 segundos.)

```
SELECT * FROM `dados` WHERE `reg_ans` NOT IN (SELECT relatorio.registroans FROM relatorio) GROUP BY `reg_ans`;
```

☐ Perfil

[Editar em linha] [Editar] [Explicar SQL] [Criar código PHP] [Actualizar]

☐ Mostrar tudo

Número de registros: 100

Filtrar registros:

Ordenar pela chave: Nenhum

Opções extra

data	reg_ans	cd_conta_contabil	descricao	vl_saldo_inicial	vl_saldo_final
2020-01-01	300012	4639	OUTRAS DESPESAS COM LOCALIZAÇÃO E FUNCIONAMENTO	0	12428
2020-01-01	302490	211111012	Provisão de Prêmio/Contraprestação Não Ganha - Pla...	0	0
2020-01-01	306444	123	CRÉDITOS DE OPERAÇÕES COM PLANOS DE ASSISTÊNCIA À ...	0	177105
2020-01-01	306959	332119212	(-) Dedução de Receitas com Prestação de Serviços	0	-3808
2020-01-01	308005	1	ATIVO	0	15771801
2020-10-01	308277	1	ATIVO	0	291397
2020-01-01	309699	1	ATIVO	0	5936439
2020-01-01	310361	4619	OUTRAS DESPESAS COM PESSOAL PRÓPRIO	0	3354
2021-01-01	311405	1	ATIVO	0	38078095
2020-01-01	313904	1	ATIVO	0	40097098
2020-01-01	314170	4589	DESPESAS FINANCEIRAS DIVERSAS	0	827
2020-01-01	316741	1	ATIVO	0	26264102
2020-10-01	323632	1	ATIVO	0	422784
2020-01-01	328598	216219	Retenções de Imposto e Contribuições	0	237334
2020-01-01	332291	123422011	Contraprestação Corresponsabilidade Assumida	0	0
2020-01-01	333328	133219034	Móveis e utensílios	0	-286505
2020-01-01	336165	3119	(-)OUTRAS DEDUÇÕES DAS CONTRAPRESTAÇÕES / PRÊMIOS	0	-1522

Importei **todos** os arquivos.csv referentes aos últimos 2 anos encontrados no repositório público na tabela dados. Foram realizados 5.965.857 registros na tabela.

Tabela	Ação	Registos	Tipo	Agrupamento (Collation)	Tamanho	Suspensão
<input type="checkbox"/> dados	Procurar Estrutura Pesquisar Inserir Limpa Eliminar	5,965,857	InnoDB	utf8_general_ci	737.6 MB	-
<input type="checkbox"/> relatorio	Procurar Estrutura Pesquisar Inserir Limpa Eliminar	1,155	InnoDB	utf8_general_ci	368.0 KB	-
2 tabela(s) Soma		5,967,012	InnoDB	utf8_general_ci	738 MB	0 Bytes

Assim é possível verificar que, para 5.965.857 registros, foram identificados 70 registros que não possuem vínculo com a tabela “relatorio”.