

Universidade da Beira Interior

Departamento de Informática



Departamento de
Informática

Desempenho de Bases de Dados

Tecnologias de Bases de Dados

Elaborado por:

João Brito nº M9984

Luís Pereira nº M10156

Carlos Esteves nº E10304

Professor:

Rui Cardoso

Covilhã, Dezembro de 2019

Acrónimos

BD *Base de Dados*

SQL *Structured Query Language*

SGBD *Sistema de Gestão de Bases de Dados*

Capítulo 1

Introdução

1.1 Enquadramento

O presente documento enquadra-se na Unidade Curricular de Tecnologias de Bases de Dados, inserida no 1º ano do Mestrado em Engenharia Informática da Universidade da Beira Interior, no ano letivo 2019/2020.

1.2 Motivação

Este trabalho coloca em prática, e aprofunda, os conhecimentos adquiridos nas aulas teóricas e práticas do presente ano letivo.

1.3 Objetivos

Este trabalho propõe-se a usar índices com vista a diminuir os tempos despendidos na execução de *queries*. Serão estudados e implementados os vários tipos de índices suportados pelo *Sistema de Gestão de Bases de Dados* (SGBD) usado (PostgreSQL).

1.4 Organização do Documento

De modo a refletir o trabalho que foi feito, este documento encontra-se estruturado da seguinte forma:

1. O primeiro capítulo – **Introdução** – apresenta o enquadramento, motivação e objetivos do projeto, bem como a organização do presente documento.

2. O segundo capítulo – **Modelo de Dados** – descreve a estrutura e detalhes de implementação do modelo de dados empregue,
3. O terceiro capítulo – **SGBD e esquema relacional** – apresenta o primeiro grande objetivo deste projeto, o método explorado na sua concretização e os resultados obtidos.
4. O quarto capítulo – **Testes** – apresenta, os testes efetuados, nomeadamente, a análise de tempos de execução com e sem índices.
5. O quinto capítulo – **Conclusão e Trabalho Futuro** – resume os conhecimentos e aptidões obtidos através da realização deste trabalho, realçando, ainda, aspetos que poderiam ser melhorados e/ou acrescentados.

Capítulo 2

Modelo de Dados

2.1 Introdução

O modelo de dados utilizado neste trabalho consiste numa tabela única com dados sobre acidentes.

2.2 Modelo de Dados

A opção tomada foi a de agregar todos os atributos dos acidentes (data, natureza, meios envolvidos, entre outros) numa tabela, sendo que, a nível de ganhos de desempenho, várias estruturas auxiliares (índices) foram utilizadas.

O modelo de dados apresenta-se de seguida:

Acidente
<ul style="list-style-type: none">- IdAcidente <small>bigserial PRIMARY KEY</small>- DataOcorrencia <small>char(40)</small>- DataFecho <small>char(40)</small>- Natureza <small>char(200)</small>- EstadoOcorrencia <small>char(200)</small>- Distrito <small>char(20)</small>- Concelho <small>char(20)</small>- Freguesia <small>char(500)</small>- Localidade <small>char(500)</small>- Latitude <small>double precision</small>- Longitude <small>double precision</small>- NumMeiosTerrestres <small>integer</small>- NumOperacionaisTerrestres <small>integer</small>- NumMeiosAereos <small>integer</small>- NumOperacionaisAereos <small>integer</small>

Figura 2.1: Modelo de dados

Capítulo 3

SGBD e esquema relacional

3.1 Introdução

De modo a reconstruir a *Base de Dados* (BD) desenvolvida no âmbito do presente projeto, o ficheiro "create_db.sql" é o ficheiro de referência. Adicionalmente, o ficheiro "queries.sql" contém as *queries* utilizadas na secção de testes, ao passo que o ficheiro "indexes.sql" alberga as instruções *Structured Query Language* (SQL) para criação dos índices.

3.2 Otimização do SGBD

Foi utilizada a ferramenta *online* PG Tune com vista a adequar as configurações gerais do SGBD PostgreSQL ao sistema computacional em uso. Em baixo apresenta-se um trecho de código que representa essas mesmas configurações:

```
# DB Version: 12
# OS Type: linux
# DB Type: desktop
# Total Memory (RAM): 14 GB
# CPUs num: 6
# Data Storage: ssd

max_connections = 20
shared_buffers = 896MB
effective_cache_size = 3584MB
maintenance_work_mem = 896MB
checkpoint_completion_target = 0.5
wal_buffers = 16MB
default_statistics_target = 100
random_page_cost = 1.1
```

```
effective_io_concurrency = 200
work_mem = 12743kB
min_wal_size = 100MB
max_wal_size = 1GB
max_worker_processes = 6
max_parallel_workers_per_gather = 3
max_parallel_workers = 6
```

3.3 Esquema relacional

Tendo em conta o modelo de dados na secção 2.2, o esquema relacional é o que se segue:

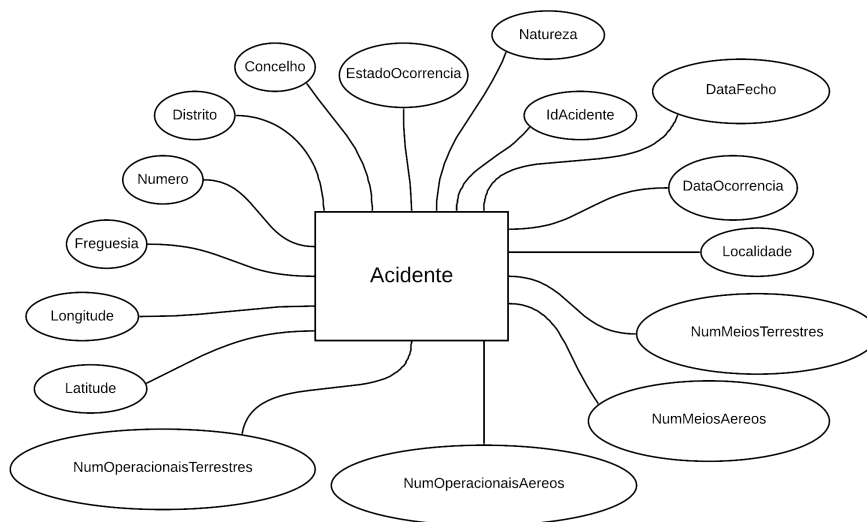


Figura 3.1: Esquema relacional

Na figura 4.1 é possível observar que existem diversos atributos na tabela de acidentes, sendo que algumas *queries* se focam, por exemplo, no IdAcidente e outras na Natureza.

Capítulo 4

Testes

4.1 Introdução

Na presente secção, serão descritos os testes efetuados, nomeadamente na exploração dos tempos de execução esperados pelo *query planner*. As *queries* foram executadas antes da implementação de qualquer índice para obter a performance base. De seguida, a mesma *query* foi implementada já com o respetivo índice criado.

Como se observará de seguida, o *query planner* optou, por vezes, pelo uso de índices, e por outras, não.

De notar que foi utilizada a combinação de comandos *EXPLAIN ANALYZE* para obtenção e futura análise de tempos (estimativas de execução, dadas pelo *query planner*).

4.2 PostGIS

Foi usada a extensão PostGIS com vista a visualizar os pontos geográficos no mapa. A imagem que se segue demonstra apenas uma pequena porção de pontos mapeados.

É possível ver a aglomeração de pontos nalgumas zonas, o que leva a concluir que existiram mais ocorrências nesses locais.

A ferramenta PostGIS foi muito útil pois permitiu dar uma perspetiva mais realista aos dados, facilitando o trabalho global.

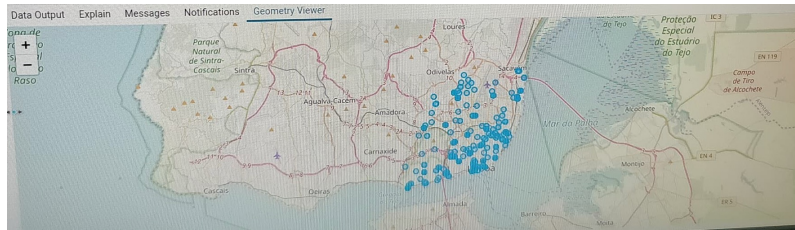


Figura 4.1: Visualização de pontos com o PostGIS

4.3 Testes com B-tree

1ª query executada:

```
EXPLAIN ANALYZE SELECT * FROM Acidentes WHERE NumMeiosTerrestres
    > 1 AND NumMeiosTerrestres < 4 AND NumMeiosAereos > 1 AND
    NumMeiosAereos < 3;
```

Índice criado:

```
CREATE INDEX idx_acidentes_meios_terrestres_btree ON Acidentes (
    NumMeiosTerrestres , NumMeiosAereos );
```

Tempo antes	Tempo depois	Delta
377.503 ms	32.953 ms	344.55 ms (11.4x)

Tabela 4.1: Testes da primeira *query*

2ª query executada:

```
EXPLAIN ANALYZE SELECT * FROM Acidentes WHERE
    NumOperacionaisTerrestres < 3;
```

Índice criado:

```
CREATE INDEX idx_acidentes_operacionais_terrestres_btree ON
    Acidentes USING btree (NumOperacionaisTerrestres );
```

Tempo antes	Tempo depois	Delta
700.542 ms	720.298 ms	- 19.756 ms (0.97x)

Tabela 4.2: Testes da segunda *query*

4.4 Testes com Hash

1ª query executada:

```
EXPLAIN ANALYZE SELECT * FROM Acidentes WHERE IdAcidente = 406700;
```

Índice criado:

```
CREATE INDEX idx_acidentes_id_hash ON Acidentes USING hash (IdAcidente);
```

Tempo antes	Tempo depois	Delta
0.272 ms	0.265 ms	0.007 ms (1.02x)

Tabela 4.3: Testes da primeira *query*

2ª query executada:

```
EXPLAIN ANALYZE SELECT * FROM Acidentes WHERE Concelho = 'covilha';
```

Índice criado:

```
CREATE INDEX idx_acidentes_concelhos_hash ON Acidentes USING hash (Concelho);
```

Tempo antes	Tempo depois	Delta
532.049 ms	0.143 ms	531.906 ms (3720.6x)

Tabela 4.4: Testes da segunda *query*

4.5 Testes com BRIN

1ª query executada:

```
EXPLAIN ANALYZE SELECT Natureza FROM Acidentes WHERE IdAcidente > 100 and IdAcidente < 20000;
```

Índice criado:

```
CREATE INDEX idx_acidentes_natureza_brin ON Acidentes USING brin  
(Natureza) WHERE IdAcidente > 100 and IdAcidente < 20000;
```

Tempo antes	Tempo depois	Delta
31.181 ms	28.191 ms	2.99 ms (1.10x)

Tabela 4.5: Testes da primeira *query*

2ª query executada:

```
EXPLAIN ANALYZE SELECT IdAcidente FROM Acidentes WHERE  
NumOperacionaisTerrestres < 2;
```

Índice criado:

```
CREATE INDEX idx_acidentes_id_brin ON Acidentes USING brin(  
IdAcidente);
```

Tempo antes	Tempo depois	Delta
240.766 ms	234.30 ms	6.46 ms (1.02x)

Tabela 4.6: Testes da segunda *query*

Capítulo 5

Conclusões e Trabalho Futuro

5.1 Conclusões Principais

Este trabalho serviu, acima de tudo, como uma oportunidade para implementar o mecanismo de indexação numa BD, com o intuito de melhorar a sua performance. Foi possível ver, na prática, como os índices diminuem os tempos de execução, pelo que o objetivo principal foi atingido.

5.2 Trabalho Futuro

De forma a suplementar o trabalho desenvolvido, algumas melhorias/adições podem ser destacadas:

1. Explorar outro tipo de índices;
2. Utilizar uma BD massiva, ainda maior que a utilizada no presente projeto.

