

Bases de Dados 2019/2020

Enunciado do projecto

Parte 4

O quarto projeto da disciplina de Bases de Dados consiste na optimização da base de dados do *Sistema TRANSLATERIGHT* no SGBD POSTGRES, implementação de transacções e implementação de um Data Warehouse.

Restrições de Integridade

Escreva o código que implemente as restrições de integridade RI-1, RI-4, RI-5 e RI-6 da Parte 3 do projecto com as extensões procedimentais do SQL (*Stored Procedures* e *Triggers*) no esquema que elaborou na Parte 3.

Índices

Baseando-se no modelo relacional do enunciado da Parte 3, crie o(s) índice(s) em SQL que considerar necessários para cada um dos casos indicados em seguida.

Suponha que não existe qualquer índice sobre as tabelas para além dos implícitos na declaração das chaves primárias e estrangeiras. Se for necessário, pode alterar a ordem dos campos das chaves primárias na declaração da tabela para que o índice associado à chave primária sirva também para otimizar a query pretendida.

Suponha que a dimensão das tabelas ultrapassa em várias ordens de grandeza a memória disponível. Suponha ainda que os blocos do disco são de 4KB.

Indique, justificando, que tipo de índice(s), sobre que atributo(s) e sobre que tabela(s) faria sentido criar de modo a acelerar a execução de cada uma das interrogações.

Justifique as suas escolhas e pressupostos.

1. Listar todos as propostas de correção entre dois momentos no tempo:

```
select email, nro, texto
from proposta_de_correcao
where data_hora between <data_hora_i> and <data_hora_f>
```

Suponha que a dimensão média do atributo `texto` é de 0.125Kb.

Proponha índice(s) para cada um dos cenários descritos abaixo.

1.1 - Em 80% das invocações, a interrogação acima devolve mais do que 10% do total de registos da tabela.

1.2 - Em 80% das invocações, a interrogação acima devolve menos de 0.001% do total de registos da tabela.

2. Devolver o email dado o ID de uma incidência:

```
select anomalia_id, email
from incidencia
where anomalia_id = some_value
```

3. Listar todos os emails de todas as correções com um `anomalid_id` superior a um determinado valor, para cada um dos seguintes cenários:

```
select email
from correcao
where anomalid_id > <some_value>
```

Proponha índice(s) para cada um dos cenários descritos abaixo.

3.1 - Em 80% das invocações, a interrogação acima devolve mais do que 10% do total de registos da tabela.

3.2 - Em 80% das invocações, a interrogação acima devolve menos de 0.001% do total de registos da tabela.

4. Listar todas as línguas e descrições que têm a flag `tem_anomalia_redacao` a `True`, entre dois momentos no tempo e estão descritas num conjunto de línguas determinadas através do operador `LIKE` contra um padrão prefixo na forma `<SOME_PATTERN>%`.

```
select id, lingua, descricao
from anomalia
where ts between <ts1> and <ts2>
and tem_anomalia_redacao = True
and language like '<SOME_PATTERN>%'
```

Modelo Multidimensional

Crie na base de dados um esquema em estrela com informação relativa a itens e anomalias tendo como dimensões:

d_utilizador(id_utilizador, tipo)

d_tempo(id_tempo, dia, dia_da_semana, semana, mes, trimestre, ano)

d_local(id_local, latitude, longitude, nome)

d_lingua(id_lingua, lingua, país)

f_anomalia(id_utilizador, id_tempo, id_local, id_lingua, tipo_anomalia, com_proposta)

- id_utilizador: FK(d_utilizador)
- id_tempo: FK(d_tempo)
- id_local: FK (d_local)
- id_lingua: FK(d_lingua)

Escreva as instruções SQL necessárias para definir e carregar o esquema em estrela a partir das tabelas existentes, carregando primeiro as tabelas de dimensões e depois a tabela de factos. Os atributos **id_utilizador**, **id_tempo**, **id_local** e **id_lingua** são chaves substitutas.

Data Analytics

Considerando o esquema em estrela criado na questão anterior, escreva uma interrogação SQL que permita analisar o total de anomalias reportadas por **tipo**, **língua** e **dia_da_semana**. A solução apresentada deverá recorrer às instruções ROLLUP, CUBE, GROUPING SETS, ou à união ("UNION") de cláusulas GROUP BY.

Relatório

O projeto será avaliado a partir do relatório entregue pelos alunos e pela discussão. O relatório deverá conter todas as respostas aos itens pedidos acima. Na tabela seguinte indica-se a valorização de cada parte do trabalho a desenvolver.

Item	Valores
Restrições de integridade	6
Índices	6
Modelo Multidimensional	3
Data analytics	5

O relatório deverá começar com uma folha de rosto com a indicação "**Projeto de Bases de Dados, Parte 4**", o **nome e número dos alunos**, a percentagem relativa de contribuição de cada um, juntamente com o esforço (em horas) que cada elemento do grupo dedicou ao projeto, o **número do grupo**, o **turno** a que o grupo pertence, o **nome do docente de laboratório** e, além da folha de rosto, o relatório deverá ter no máximo **6 páginas**.

O trabalho terá que ser entregue em duas versões:

1. **Versão digital**, em formato ZIP, a entregar via Fénix até às 23h59 da data de entrega.
2. **Versão em papel**, a trazer em mão para a discussão.