

O Trabalho 1 envolve os conceitos de SQL, Indexação e Processamento de consultas. Bem como, a construção e o povoamento da base de dados a ser utilizada.

Cada equipe deve escolher uma base de dados abertos, cujo download pode ser feito no link <http://dados.gov.br/>. Escolha uma base em que apresente um bom volume de dados, pois você irá exercitar o uso de índices na sua base de dados escolhida.

Após escolher qual base você utilizará, a partir da modelagem relacional, você deve construir o esquema do banco de dados no PostgreSQL. O seu esquema deve conter **pelo menos 2 tabelas do seu modelo relacional, portanto você pode trabalhar com parte da base escolhida. No entanto, escolha tabelas que apresentem grande quantidade de tuplas e que estejam relacionadas/ligadas umas as outras.**

Os dados podem estar dispostos em arquivos texto, csv, entre outros. Dessa forma, é necessário que você realize um processo de ETL (Extração, Transformação e carga) para que os mesmos sejam persistidos no SGBD. Você pode realizar isso via Pentaho ou utilizando o driver JDBC para inserir nas tabelas do seu modelo. Essa parte do trabalho fica a seu critério.

**Você deve aplicar os dois passos a seguir na sua base, após ser criada:**

- Inicialmente, verifique quais índices sua base de dados possui. Provavelmente, os associados a chaves primárias. Dessa forma, é necessário removê-los para que os resultados coletados em cada questão não sejam interferidos pela presença de outros índices.
- Remova as chaves primárias e estrangeiras usando o seguinte comando: **alter table (nome da tabela) drop constraint (nome da restrição).**

Você deve entregar na forma de um relatório as respostas, exceto a Consulta 5. Dia 16 de Maio é a entrega do trabalho em sala!!

#### Consulta 1:

Escolha uma tabela da sua base de dados. Certifique-se de que nela não há índices. Defina uma consulta do tipo *exact-match* utilizando apenas essa tabela escolhida, projetando 3 atributos que pertencem a mesma e sendo a condição de seleção uma igualdade e sobre um atributo apenas. Escolha um valor apropriado 'xxxxx'. No exemplo, imagine que eu escolhi a tabela A.

```
SELECT A.atributo1, A.atributo2, A.atributo3  
  
FROM tabela A  
  
WHERE A.atributo4='xxxxx'
```

- a) Em quanto tempo você executou sua consulta?
- b) Crie um índice do tipo Hash sobre o atributo que você utilizou na sua condição de seleção. Em quanto tempo o índice foi criado?
- c) Limpe a memória cache do sistema operacional e do SGBD que você está utilizando. Agora execute a mesma consulta e informe em quanto tempo você obteve o resultado. Certifique-se que sua consulta é executada utilizando o índice criado por você na letra b, para isso utilize o plano de consulta (em geral, comando “Explain Analyse”).
- d) Exclua o índice Hash criado. Crie um índice do tipo BTREE sobre o atributo que você utilizou na sua condição de seleção. Em quanto tempo o índice foi criado?
- e) Limpe a memória cache do sistema operacional e do SGBD que você está utilizando. Agora execute a mesma consulta e informe em quanto tempo você obteve o resultado. Certifique-se que sua consulta é executada utilizando o índice criado por você na letra d, para isso utilize o plano de consulta (em geral, comando “Explain”).
- f) O que você pode afirmar ao comparar as respostas dos itens “c” e “e”?

#### Consulta 2:

Escolha uma tabela da sua base de dados. Certifique-se de que nela não há índices. Defina uma consulta do tipo *select-range* utilizando apenas essa tabela escolhida, projetando 3 atributos que pertencem a mesma e sendo a condição de seleção uma operação de “maior que” sobre um atributo apenas. No exemplo, imagine que eu escolhi a tabela A. Escolha um valor apropriado para ‘xxxxx’.

```
SELECT A.atributo1, A.atributo2, A.atributo3  
  
FROM tabela A  
  
WHERE A.atributo4 > 'xxxxx'
```

- a) Em quanto tempo você executou sua consulta?
- b) Crie um índice do tipo BTREE sobre o atributo que você utilizou na sua condição de seleção. Em quanto tempo o índice foi criado?
- c) Limpe a memória cache do sistema operacional e do SGBD que você está utilizando. Agora execute a mesma consulta e informe em quanto tempo você obteve o resultado. Certifique-se que sua consulta é executada utilizando o índice criado por você na letra b, para isso utilize o plano de consulta (em geral, comando “Explain Analyse”).
- d) Caso a consulta não utilize o índice, verifique e informe um valor para ‘xxxxx’ para que o SGBD use o índice.

- e) Exclua o índice BTREE criado. Crie um índice do tipo Hash sobre o atributo que você utilizou na sua condição de seleção. Em quanto tempo o índice foi criado?
- f) Limpe a memória cache do sistema operacional e do SGBD que você está utilizando. Agora execute a mesma consulta e informe em quanto tempo você obteve o resultado. Certifique-se que sua consulta é executada utilizando o índice criado por você na letra d, para isso utilize o plano de consulta (em geral, comando “Explain Analyse”).
- g) O que você pode afirmar ao comparar as respostas dos itens “c”, “d” e “e”?

### Consulta 3:

Escolha duas tabelas da sua base de dados em que uma referencie a outra via chave estrangeira. Certifique-se de que nelas não há índices. Defina uma consulta do tipo junção utilizando apenas essas tabelas escolhidas, projetando 3 atributos que pertencem a alguma delas e sendo a condição de seleção uma operação de junção sobre a chave primária de uma tabela com a chave estrangeira da outra. No exemplo, imagine que eu escolhi as tabelas A e B. Sendo A.atributo\_pk a chave primária de A e B.atributo\_fk a chave estrangeira de B que referencia a primária de A.

```
SELECT A.atributo1, B.atributo2, A.atributo3  
  
FROM tabela A, tabela B  
  
WHERE A.atributo_pk=B.atributo_fk
```

- a) Em quanto tempo você executou sua consulta?
- b) Crie um índice do tipo Hash sobre A.atributo\_pk.
- c) Limpe a memória cache do sistema operacional e do SGBD que você está utilizando. Agora execute a mesma consulta e informe em quanto tempo você obteve o resultado. Certifique-se que sua consulta é executada utilizando o índice criado por você na letra b, para isso utilize o plano de consulta (em geral, comando “Explain”).
- d) Crie um índice tipo Hash sobre o atributo que B.atributo\_fk
- e) Limpe a memória cache do sistema operacional e do SGBD que você está utilizando. Agora execute a mesma consulta e informe em quanto tempo você obteve o resultado. Certifique-se que sua consulta é executada utilizando o índice criado por você na letra d, para isso utilize o plano de consulta (em geral, comando “Explain”).
- f) O que você pode afirmar ao comparar as respostas dos itens “c” e “e”?

#### Consulta 4:

Escolha uma tabela da sua base de dados. Certifique-se de que nela não há índices. Defina uma consulta do tipo agregação utilizando apenas essa tabela escolhida, projetando 1 atributo que pertence a mesma e a função de agregação. No exemplo, imagine que eu escolhi a tabela A.

```
SELECT A.atributo1, COUNT(*)  
  
FROM tabela A  
  
GROUP BY A.atributo1
```

- Em quanto tempo você executou sua consulta?
- Crie um índice do tipo Hash sobre A.atributo1. Em quanto tempo o índice foi criado?
- Limpe a memória cache do sistema operacional e do SGBD que você está utilizando. Agora execute a mesma consulta e informe em quanto tempo você obteve o resultado. Certifique-se que sua consulta é executada utilizando o índice criado por você na letra b, para isso utilize o plano de consulta (em geral, comando “Explain”).

#### Consulta 5:

Crie uma consulta diferente das anteriores em que você utilize apenas uma tabela. Certifique-se que nessa tabela não há índices.

Nesta consulta, aplique a condição de seleção sobre um atributo que não é chave primária da tabela. Lembre-se de antes de executar a consulta: Limpar o cache!!.

Você deve fazer uma pequena aplicação usando JDBC e o padrão DAO para reportar o tempo de execução das consultas com e sem o índice. O índice é você quem deve sugerir a fim de acelerar o processamento da consulta. Bem como, na sua aplicação deve permitir ao usuário verificar o plano de consulta.

Bom trabalho!!