

- O Trabalho 1 envolve os conceitos de SQL, Indexação e Processamento de Consultas.
- O Trabalho 1 vale 30% da nota da AV1 e deve ser feito em dupla.
- A data de entrega do trabalho é dia 13/05/2019. Você deve entregar na forma de um relatório descrevendo a sua base de dados (Parte 1) e as respostas da Parte 2.

### Parte 1 – Escolha, construção e povoamento das bases de dados.

Cada equipe deve escolher uma base de dados. Como sugestão, o download pode ser feito em um dos links abaixo. Escolha uma base em que apresente pelo menos 1 GB de dados. Quanto maior for sua base de dados, melhor será para fazer a análise. A escolha da base de dados é importante. Quanto mais interessante for a base de dados, melhor.

Alguns sites que disponibilizam dados:

- Portal de dados abertos do governo Fdral: <http://dados.gov.br/>
- Portal de dados abertos de Fortaleza: <http://dados.fortaleza.ce.gov.br/>
- Portal de ddos abertos do Ceará: [https://cearatransparente.ce.gov.br/portal-da-transparencia/dados-abertos/conjuntos-de-dados?locale=pt-BR&\\_ =](https://cearatransparente.ce.gov.br/portal-da-transparencia/dados-abertos/conjuntos-de-dados?locale=pt-BR&_=)

A escolha dos dados não está restrita a esses links.

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 4 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 umas com 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. Para o processo ETL, utilize o componente Pentaho Data Integration (PDI, também conhecida como Kettlo) .

No link abaixo você encontra informações sobre o Kettle.  
<http://community.pentaho.com/projects/data-integration/>.

### Parte 2 – Execução e Análise das Consultas

Como segunda parte do trabalho, você irá analisar o processamento das consultas com/sem estruturas de índices na base de dados finalizada por você nessa primeira parte.

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).

### 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'
```

- Em quanto tempo a consulta executou?
- Crie um índice do tipo Hash sobre o atributo que você utilizou na sua condição de seleção. 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 letra b, para isso utilize o plano de consulta (em geral, comando "Explain Analyse"). Explique o plano de consulta apresentado pelo SGBD.
- 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?
- 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 letra d, para isso utilize o plano de consulta (em geral, comando "Explain Analyse"). Explique o plano de consulta apresentado pelo SGBD.
- 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'
```

- Em quanto tempo a consulta executou?
- Crie um índice do tipo BTREE sobre o atributo que você utilizou na sua condição de seleção. 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 letra b, para isso utilize o plano de consulta (em geral, comando "Explain Analyse"). Explique o plano de consulta apresentado pelo SGBD.

- 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 letra d, para isso utilize o plano de consulta (em geral, comando "Explain Analyse"). Explique o plano de consulta apresentado pelo SGBD.
- g) O que você pode afirmar ao comparar as respostas dos itens "c", "d" e "e"?
- h) Verifique a correlação estatística entre ordem física e ordem lógica dos valores do atributo escolhido. Qual o impacto que isso teve no processamento consulta?

### 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 a consulta executou?
- 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 letra b, para isso utilize o plano de consulta (em geral, comando "Explain Analyse"). Explique o plano de consulta apresentado pelo SGBD.
- 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 Analyse"). Explique o plano de consulta apresentado pelo SGBD.
- 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
```

- a) Em quanto tempo a consulta executou?
- b) Crie um índice do tipo Hash sobre A.atributo1. 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 letra b, para isso utilize o plano de consulta (em geral, comando “Explain Analyse”). Explique o plano de consulta apresentado pelo SGBD.

### **Consulta 5:**

Exercite sua criatividade!!!

Crie duas consultas diferentes das anteriores. Certifique-se que nessa tabela não há índices. Suas consultas devem pelo menos uma junção, um comando de seleção e um operador especial (EXISTS, IN, GROUP BY, etc). Lembre-se de antes de executar a consulta: Limpar o cache e verificar se o SGBD está escolhendo utilizar o índice no plano de consulta.

Quais índices você criou para executar cada uma das consultas? Reporte o tempo de execução das consultas com e sem os índices, explique o porquê da sua escolha.

Bom trabalho!!