

ANÁLISE DE PERSISTÊNCIA DE IMAGENS MÉDICAS: UMA COMPARAÇÃO ENTRE OS SISTEMAS DE BANCOS DE DADOS *MYSQL*, *POSTGRESQL* E *DERBY*

Sérgio Roberto Delfino¹; Lucas Venezian Povoá²; Ana Carla Rossinholi Pinto³

Resumo

Este artigo apresenta a análise e comparação de desempenho de persistência de imagens médicas de três Sistemas Gerenciadores de Banco de Dados Open Source, MySQL 5.1, PostgreSQL 9.0 e Derby 10.8, através de medidas de posição e dispersão amostral, histogramas e do Teste de Tukey.

Palavras-chave: Desempenho, Teste de Tukey, Sistemas Gerenciadores de Banco de Dados

Abstract

This paper presents the performance analysis and comparisons of medical images persistence among three Open Source Database Management System, MySQL 5.1, PostgreSQL 9.0 and Derby 10.8, through dispersion and position measures, histograms and Tukey Test.

Key words: Performance, Tukey Test, Database Management Systems

1 INTRODUÇÃO

O tratamento de dados com grande massa de informação é um desafio tecnológico atual, havendo esse tipo de situação em diversos setores, tais como educação, geoprocessamento, *marketing* e saúde. Um dos casos particulares desse contexto é a manipulação de imagens médicas, as quais possuem características específicas, devendo haver um tratamento particular para esse tipo de dado. Portanto, esse artigo apresenta um estudo de desempenho feito sobre três Sistemas Gerenciadores de Banco de Dados (SGBD) *Open Source* com espaço significativo no mercado atual, sendo eles: MySQL 5.1 (Oracle, 2011b), PostgreSQL 9.0 (PostgreSQL Global Development Group, 2011) e Derby 10.8 (The Apache DB Project, 2011). Dessa forma, a Seção 2 apresenta a metodologia utilizada para analisar os Sistemas de Banco de Dados, bem como uma ferramenta desenvolvida com o propósito de obter os tempos de persistência de cada Banco de Dados. A Seção 3 trata das análises e discussões sobre os resultados. A Seção 4 descreve trabalhos a serem desenvolvidos na mesma linha de pesquisa. A Seção 5 apresenta as considerações finais. Por fim, a última Seção lista as referências utilizadas neste artigo.

¹ Faculdade de Tecnologia de Ourinhos (FATEC) e Faculdades Integradas de Ourinhos (FIO). Email: srdelfino@gmail.com.

² Departamento de Computação (DC) da Universidade Federal de São Carlos (UFSCar). Email: veneziam@ourinhos.unesp.br.

³ Graduada na Faculdade de Tecnologia de Ourinhos (FATEC). Email: carlarossinholi@gmail.com.

2 FERRAMENTAS E MÉTODOS

A metodologia para o desenvolvimento deste estudo é composta por duas etapas: a primeira é a coleta dos tempos de persistência de um conjunto de 2486 imagens médicas, todas com o tamanho de 128 MiB, efetuada pelo software BSMIP⁴ (maiores detalhes sobre esse software serão expostos na Seção 3.1), desenvolvido especificamente para essa finalidade.

O sistema para coleta dos tempos de persistência, como o Sistema Gerenciador de Banco de Dados, foram executados em um mesmo *host* com as seguintes características físicas: 2 GiB de Memória RAM, processador Intel Pentium(R) Dual CPU T2390 com 1,86 GHz de frequência e HD de 200 GiB. O Sistema Operacional utilizado foi o Debian 6.0.3 (SPI and other, 2011), Kernel Linux 2.6.32-5-686. A execução dos Sistemas de Banco de Dados foi feita isolada e independentemente e sob as mesmas condições, não havendo concorrência com outros processos, exceto os que são requisitos mínimos do Sistema Operacional.

A segunda etapa foi a análise dos tempos de persistência, efetuada através de medidas de posição e dispersão amostral, da exposição dos histograma das amostras de cada Sistema de Banco de Dados e também pelo Teste de Tukey para verificar se há diferença estatisticamente significantes entre as médias dos três Sistemas sendo que tais análises foram efetuadas utilizando-se recursos do Ambiente e Linguagem R⁵ (R Core Team, 2011).

As imagens médicas foram persistidas nas tabelas com as estruturas descritas pelo Quadro 1.

Quadro 1 - Estruturas utilizadas para persistência de imagens médicas nos Sistemas de Banco de Dados a) MySQL, b) PostgreSQL e c) Derby

a)	b)	c)
CREATE SCHEMA	CREATE SCHEMA	CREATE SCHEMA
BSMIP;	BSMIP;	BSMIP;
CREATE TABLE Bsmip	CREATE TABLE Bsmip	CREATE TABLE Bsmip
(((
image	image BYTEA	image BLOB
LARGEBLOB););
);		

Fonte: Elaborado pelos autores

⁴ O código fonte do software BSMIP encontra-se disponível em <http://www.lucasvenezian.orgfree.com/projetos/bsmip/bsmip-1.0-src.tar.gz>

⁵ O código fonte, bem como os dados utilizados para análise dos Sistemas de Banco de Dados, encontram-se disponíveis em <http://www.lucasvenezian.orgfree.com/projetos/bsmip/bsmip-analysis.tar.gz>

2.1 Benchmark System for Medical Images Persistence

O software Benchmark System for Medical Images Persistence (BSMIP) foi desenvolvido sob o paradigma Orientado a Aspectos, sendo utilizado para isso as tecnologias Java 6 (Oracle, 2011a) e AspectJ 1.6.11 (Eclipse Team, 2011). Sua finalidade é fornecer uma interface simples ao usuário para efetuar persistência de imagens ou arquivos binários em um determinado Sistema Gerenciador de Banco de Dados, especificando também o esquema, a tabela e a coluna onde se deseja gravar o conjunto de imagens, e também fornecer como resultado um arquivo XML com os seguintes dados: quantidade de imagens persistidas, nome do SGBD analisado e para cada imagem gravada o seu endereço absoluto, tempo para ser persistida em nanosegundos e seu tamanho em *bytes*. A estrutura do arquivo XML emitido como resposta pelo BSMIP é descrito pelo Quadro 2.

Quadro 2 - Parte de um arquivo XML emitido pelo BSMIP

```
<?xml version="1.0" encoding="UTF-8"?>

<!DOCTYPE persistence-benchmark
PUBLIC "-//FATECOU//DTD Persistence Benchmark"
"http://www.fatecou.edu.br/dtd/persistence-
benchmark.dtd">

<persistence-benchmark>
  <dbms>postgresql</dbms>
  <number-of-images>2486</number-of-images>
  <times-of-persistence>
    <image>
      <name>/home/bsmip/images/MS10002034.TIF</name>
      <time>603840428</time>
      <bytes>131072</bytes>
    </image>
    <image>
      <name>/home/bsmip/images/MS10002035.TIF</name>
      <time>512432210</time>
      <bytes>131072</bytes>
    </image>
    <!-- Dados omitidos -->
  </times-of-persistence>
</persistence-benchmark>
```

Fonte: Elaborada pelos autores

Os tempos são marcados somente em relação a inserção das imagens, não levando em consideração outros processos, como conexão com o Sistema de Banco de Dados e início e término de transação. A Application Programming Interface (API) Java Database Connectivity (JDBC) é o único recurso utilizado para interface com o Banco de Dados, minimizando assim, a interferência das tecnologias utilizadas no processo de comunicação com o SGBD.

A Figura 1 ilustra o Diagrama de Classes do software BSMIP segundo a linguagem de modelagem UML 2.0 (BOOCH; RUMBAUGH,; JACOBSON, 2005).

possuindo tempo concentrados entre 0 e 4500 milissegundos, sendo o pior Sistemas entre os três comparados.

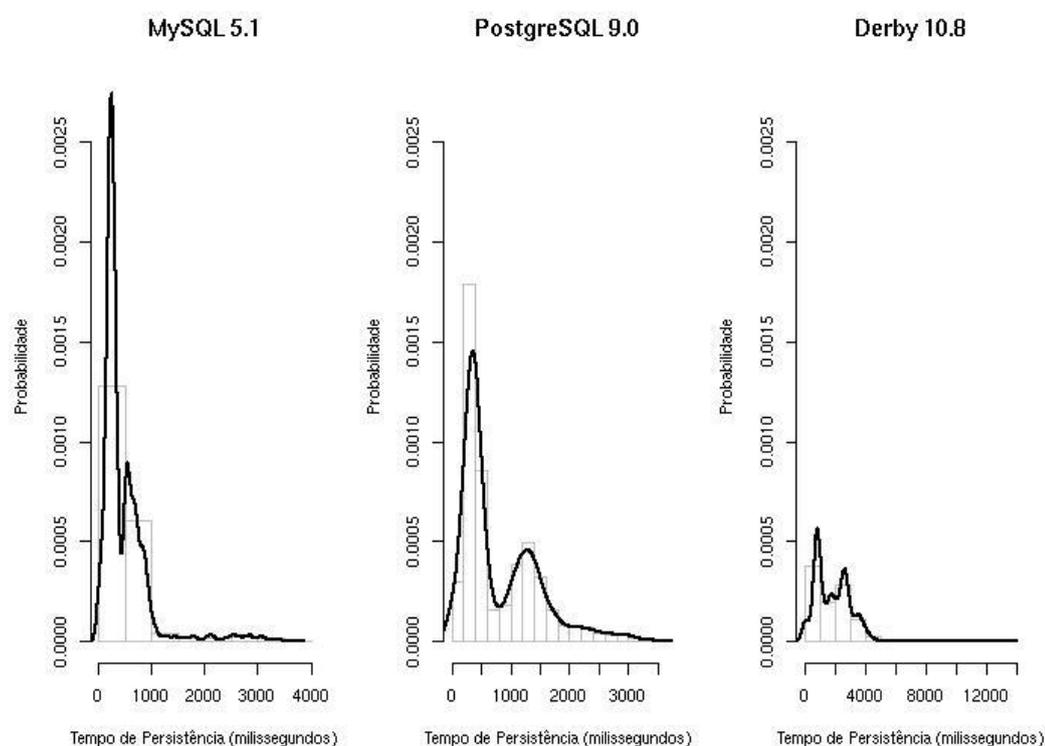


Figura 2 - Tempos de Persistência em Relação a sua Probabilidade na amostra
Fonte: Elaborada pelos autores

O Teste de Tukey é utilizado para comparar várias médias entre si, sendo extremamente útil para esse contexto, já que é necessário comparar os Sistemas de Banco de Dados par a par. O resultado dessa ferramenta estatística demonstra que há diferença significativamente estatística entre as médias dos três SGBDs, uma vez que não há nenhum valor zero nos valores do Menor Ponto Extremo e Maior Ponto Extremo, além disso o p-valor possui um dado menor do que o nível de significância de 5%, utilizado em nosso estudo, rejeitando assim a hipótese nula, ou seja a hipótese de que as médias são iguais. Os resultados do Teste de Tukey demonstram que os Sistemas MySQL e PostgreSQL não possuem diferenças tão gritantes ou que não ocorre com o Sistema Derby quando comparados com os outros sistemas. A Tabela 2 expõe os resultados do Teste de Tukey.

Tabela 1 - Resultados do Teste de Tukey com Nível de Significância de 5%

	Diferença	Menor ponto extremo	Maior ponto extremo	P-Valor
PgSQL/MySQL	311.922	256.1268	367.7173	0.0000
PgSQL/Derby	1053.536	1309.6626	1421.2531	0.0000
MySQL/Derby	1365.458	454.9000	1724.0000	0.0000

4 TRABALHOS FUTUROS

Pretende-se desenvolver o mesmo tipo de análise sobre os mesmo Sistemas de Banco de Dados, porém, aplicando a análise sobre outras formas de manipulação de imagens médicas, como a consulta, atualização e exclusão, além de utilizar um sistema mais parecido com o real, havendo intervalos aleatórios para a manipulação da imagem além de diversos clientes efetuando essas ações concorrentemente, verificando diversos aspectos comportamentais dos Sistemas de Banco de Dados. Sobre essa nova metodologia também espera-se utilizar um recurso estatístico mais sofisticado, como a Teoria de Filas.

5 CONCLUSÕES

As análises demonstram que o Sistema Gerenciador de Banco de Dados que possui melhor desempenho na persistência de imagens médicas é o MySQL, com média de tempo de 476.215 milissegundos e desvio-padrão de 484.4991, além de possuir diferença estatisticamente significantes segundo o Teste de Tukey quando comparado com o PostgreSQL e Derby. A diferença de desempenho entre o PostgreSQL e o MySQL é de aproximadamente 39,58% segundo o Teste de Tukey e entre o Derby e o MySQL5.1 é de aproximadamente 74,14%, o que permite inferir que a diferença entre os SGBD MySQL e PostgreSQL não são tão gritantes, apesar de significativas, contudo, também é possível notar através do coeficiente de variância que o PostgreSQL possui maior estabilidade em relação ao MySQL. O SGBD Derby possui notável diferença entre os outros dois sistemas, possuindo tempo médio de persistência de 1841.67 e coeficiente de variação de 65,69%, demonstrando assim maior acúmulo dos valores da amostra à média. Destarte, é possível afirmar que o Sistema de Banco de Dados entre o três analisados é o MySQL, seguido do PostgreSQL e por último o Derby.

REFERÊNCIAS

BOOCH, G.; RUMBAUGH, J.; JACOBSON, I. (2005). **UML: guia do usuário**. 2ª ed. Rio de Janeiro. Elsevier.

Oracle (2011a). **Java**. Disponível em: <<http://java.com>>. Acesso: 18 out. 2011.

Oracle (2011b). **MySQL: the world's most popular open source database**. Disponível em: <<http://www.mysql.com>>. Acesso: 20 out. 2011.

PostgreSQL Global Development Group (2011). **PostgreSQL: the world's most advanced open source database**. Disponível em: <<http://www.postgresql.org>>. Acesso: 15 out. 2011.

R Development Core Team (2011). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>. Acesso: 28 set. 2011.

SPI and other (2011). **Debian: the universal operational system**. Disponível em: <<http://debian.org>>. Acesso: 20 set. 2011.

The Apache DB Project (2011). **Apache Derby**. Disponível em: <<http://db.apache.org/derby>>. Acesso: 10 out. 2011.

The Eclipse Foundation (2011). **AspectJ: crosscutting objects for better modularity**. Disponível em: <<http://www.eclipse.org/aspectj>>. Acesso: 01 out. 2011.