GESTÃO E ANÁLISE DE DADOS DE SENSORES EM BANCO DE DADOS NÃO-RELACIONAL

João Gabriel Corrêa Krüger (BIC/Fundação Araucária/UEPG), Maria Salete Marcon Gomes Vaz (Orientador), e-mail: salete@uepg.br

Universidade Estadual de Ponta Grossa/Departamento de Informática.

Ciências Exatas e da Terra, Banco de Dados.

Palavras-chave: Comparação de desempenho, MariaDB, MongoDB, Gestão de dados.

Resumo

O presente trabalho consiste da avaliação de performance comparativa do Sistema Gerenciador de Banco de Dados (SGBD) não relacional *MongoDB* com relação ao SGBD relacional *MariaDB*. A metodologia de trabalho adotada é a de um modelo hipotético de dados de sensores e realizar teste de operações CRUD (*Create, Read, Update and Delete*) nos SGBDs em questão. É avaliado o tempo de execução das operações com diferentes tamanhos de entrada, onde se comparam os valores médios usando de testes estatísticos. Pode-se concluir que existe superioridade por parte do *MongoDB* quanto às operações *Create*, *Update* e *Delete* e inferioridade quanto à operações *Read* em relação ao *MariaDB*.

Introdução

Com o advento da Internet das Coisas, onde aplicações e eletrônicos armazenam dados sobre sua utilização para posterior uso, é crescente a necessidade por banco de dados com a capacidade de armazenar grandes quantidades de dados.

O grande influxo de dados sendo levantados por sensores e aplicações faz relevante a preocupação com a escalabilidade de banco de dados. Novos Sistemas Gerenciadores de Banco de Dados, ou SGBD, vêm sido desenvolvidos com essa preocupação em mente.

O *MongoDB* é um SGBD que conta com um banco de dados não relacional. O *MongoDB* conta com o armazenamento das entradas no formato BSON para armazenar os dados, diferindo assim de bancos de dados relacionais, visto que estes armazenam dados em um formato similar a tabelas e possuem grande facilidade de indexação.

O *MariaDB*, por sua vez, é um SGBD que conta com um banco de dados estruturado de forma relacional. Um banco de dados relacional organiza seus dados em formato de tabelas para serem consultadas com SQL, ou *Structured Query Language*.

Com o crescimento do movimento No-SQL, vêm-se estudando situações onde o desempenho dos novos bancos de dados é melhor do que os já utilizados pela comunidade acadêmica e pelo mercado (BONNET et al., 2001) (POLITOWSKI e MARAN, 2014).

Faz-se necessário, assim, realizar mais estudos comparativos entre esses bancos para identificar aplicações onde são mais adequados os usos de SGBDs não relacionais.

Com esse projeto de pesquisa pretende-se realizar de desempenho de banco de dados entre os SGBDs *MongoDB* e *MariaDB*.

Materiais e Métodos

Bancos de dados não relacionais (POLITOWSKI e MARAN, 2014) são bancos de dados que não usam o esquema de tabela e linguagem SQL para descrever requisições e dados no banco de dados. Sua origem veio da necessidade por maior escalabilidade, problema muito comum em banco de dados relacionais. Já os sensores são dispositivos que transformam grandezas físicas, com o propósito de realizar medições.

O *MongoDB* representa os dados como documentos armazenados em BSON, uma representação binária do JSON (*Javascript Object Notation*) e permite mudanças sem grandes alterações nos arquivos, o que possibilita uma maior flexibilidade e uma maior confiança no ponto de vista de disponibilidade. Possui como fragilidade a maior dificuldade de definições de regras para campos no banco de dados e o tamanho alocado em disco do banco de dados, que tende a ser maior do que um banco de dados relacional.

Politowski e Maran (2014) compararam o desempenho dos Sistemas Gerenciadores de Banco de Dados *PostgreSQL* e *MongoDB* e afirmaram a superioridade desse último, inerente aos tipos de requisições e volumes de dados no conjunto de testes desenvolvidos pelos autores.

Similar ao *PostgreSQL*, o *MariaDB* é Sistema Gerenciador de Banco de Dados relacional de código aberto. O *MariaDB* surgiu como um *fork* do *MySQL* e é um dos SGBDs mais adotados na atualidade.

Foram desenvolvidos dois programas na linguagem de programação Python, um para cada Sistema Gerenciador de Banco de Dados. Os programas realizam uma série de operações CRUD, ou *Create Read Update and Delete,* nos bancos de dados, medindo seus tempos de execução.

Foram geradas 1.000.000 entradas aleatórias, simulando dados de sensores de temperatura, luminosidade e umidade. Os dados foram modelados a partir de dispositivos em salas hipotéticas onde a temperatura, luminosidade e umidade variam de acordo com uma distribuição normal. O *range* de valores gerados não afeta o desempenho dos bancos de dados testados visto que tais bancos armazenam os dados binariamente em campos de tamanho fixo quando se trata de números.

Resultados e Discussão

Foram realizados conjuntos de testes diferentes. Os conjuntos de testes consistem de número de dados de 1, 10, 100, 1.000, 10.000, 100.000 e 1.000.000 dados gerados aleatoriamente, para verificar se há alteração no comportamento do tempo médio das respostar em função do número de entrada.

A Tabela 1 demonstra o tempo total para o tempo de execução dos testes.

Tabela 1 – Tempo total em segundos para as diferentes operações realizadas

	Create	*Read*	*Update*	*Delete*
MariaDB	1588,36507	285,118403	1677,22841	1656,79739
MongoDB	336,02213	351,17816	380,042465	355,560511

Tempos em segundos para 1.000.000 de operações realizadas seguidamente uma da outra.

O Sistema Gerenciador de Banco de Dados *MongoDB* apresentou melhores resultados nas operações de leitura, atualização e remoção de entradas com relação ao *MariaDB*. *MariaDB*, por sua vez, apresentou melhores resultados em operações de leituras.

Tabela 2 – Tempo médio em segundos para as diferentes operações realizadas

	Create	*Read*	*Update*	*Delete*
MariaDB	0,00158837	0,00028512	0,00167723	0,0016568
MongoDB	0,00033602	0,00035118	0,00038004	0,00035556

Tempos médios obtidos para as operações no decorrer de 1.000.000 de execuções.

Os tempos médios, apresentados na tabela 2, mantiveram-se consistentes com os tamanhos dos testes. Os resultados observados para o tempo total se apresentam também nos tempos médios, onde o SGBD *MongoDB* possui melhores tempos de inserção, atualização e remoção de entradas. O *MariaDB* possui melhores tempos de execução para operações de leitura.

Para verificar se a diferença nos tempos médios de operação é estatisticamente significante, foram realizados Testes T de *Student* com os dados obtidos de cada operação realizada. Os resultados dos valores-p obtidos nos testes podem ser atestados na Tabela 3

Tabela 3 –Testes T de *Student* para igualdade da média dos conjuntos de dados

	Create	*Read*	*Update*	*Delete*
Valor-P	0.000000754	0.048380	0.00000178	0.0003815
Diferença significativa (α = 0,05)	Sim	Sim	Sim	Sim

O parâmetro de nível de confiança adotado no teste foi α = 0,05. Devido ao fato dos valores-p serem menores que o nível de confiança, pode-se afirmar com 95% de certeza que as médias obtidas diferem entre si.

Usando também do Teste T de Student, pode-se comparar se os tempos médios do MongoDB são menores que o do SGBD *MariaDB*, conforme tabela 4.

Tabela 4 – Teste T de *Student* para tempo no *MongoDB* menor que *MariaDB*

	Create	*Read*	*Update*	*Delete*
Valor-P	0.000000377	0.9758	0.0008921	0.0001907
Diferença significativa (α = 0,05)	Sim	Não	Sim	Sim

Pode-se afirmar, com 95% de certeza, que os tempos no *MongoDB* são menores que os tempos no *MariaDB*, com exceção para os tempos de leitura. Para

operações de leitura, observando as tabelas 3 e 4, pode-se concluir que os tempos no *MariaDB* são menores que os tempos do SGBD *MongoDB*.

Conclusões

Com o desenvolvimento desse projeto de pesquisa pode-se concluir que o *MongoDB* é uma alternativa válida para aplicações cujos bancos de dados envolvam requisições como as descritas nesse trabalho, como é o caso de aplicações que envolvam armazenar dados de sensores.

Pode-se concluir que o Sistema Gerenciador de Banco de Dados *MongoDB* apresenta melhor desempenho que o *MariaDB* para operações de inserção, atualização e remoção de entradas em suas coleções. Aplicações onde essas operações são realizadas com frequência irão possuir melhor desempenho utilizando o banco de dados não relacional *MongoDB* ao invés de optarem por uma opção relacional com o MariaDB.

Pode-se verificar que o SGBD *MariaDB* tem desempenho similar, porém melhor, ao do *MongoDB* em casos de buscas simples por elementos usando o identificador.

Sendo assim, em aplicações onde operações de busca não são realizadas com demasiada frequência e operações de inserção são realizadas constantemente, como no caso de aplicações com muitos sensores, o SGBD não relacional *MongoDB* torna-se uma opção mais rápida que o SGBD relacional *MariaDB*.

Agradecimentos

Agradeço à Fundação Araucária pela bolsa para o desenvolvimento da pesquisa. Agradeço também a minha orientadora Maria Salete pela oportunidade de realizar um projeto de pesquisa e pela a orientação dada. Agradeço a Universidade Estadual de Ponta Grossa pela educação e estrutura, necessárias para o desenvolvimento deste projeto.

Referências

BONNET, Philippe; GEHRKE, Johannes; SESHADRI, Praveen. Towards sensor database systems. In: **International Conference on Mobile Data Management**. Springer, Berlin, Heidelberg, 2001. p. 3-14.

POLITOWSKI, Cristiano; MARAN, Vinıcius. **Comparação de Performance entre PostgreSQL e MongoDB**. X Escola Regional de Banco de Dados. SBC, p. 1-10, 2014.