

Universidade Federal de Pernambuco

Graduação em Ciência da Computação

Centro de Informática



SQL x NoSQL: Análise de desempenho do uso do MongoDB em relação ao uso do PostgreSQL.

Proposta de Trabalho de Conclusão de Curso

Aluno: Marcos André Pereira Martins Filho (mapmf@cin.ufpe.br)

Orientador: Fernando da Fonseca de Souza (fdfd@cin.ufpe.br)

Recife, 14 de Abril de 2014

1. Contexto

Com o advento da web 2.0 e a enorme quantidade de dados circulando na rede, algumas empresas, que encontraram limitações na forma relacional do gerenciamento de dados, se viram obrigadas a procurar por alternativas não-relacionais.

Neste contexto surge, o conceito do NoSQL (*Not Only SQL*, que quer dizer “Não apenas SQL”)[1], uma abordagem não-estruturada de gerenciamento de dados que preza desempenho, escalabilidade e disponibilidade, fazendo com que premissas como por exemplo, consistência, presente nos tradicionais bancos de dados relacionais fossem revistas.

2. Motivação

A motivação surgiu com a iniciativa de analisar o desempenho das duas abordagens de SGBD¹, visto as vantagens que a abordagem NOSQL oferece, como desempenho e escalabilidade.

O SGBD relacional escolhido para esse trabalho foi o PostgreSQL[2], o qual vem sendo utilizado nas atividades que o autor realiza no estágio. O MongoDB[3] foi o SGBD não-relacional escolhido, por ser um dos mais conhecidos, open-source e com uma abordagem orientada a documentos, que é algo bastante diferente da abordagem relacional convencional.

3. Objetivo

O objetivo deste trabalho é realizar uma análise comparativa entre SGBD relacional e não-relacional, utilizando uma base com um volume de dados considerável.

¹ SGBD: Sistema de Gerenciamento de Banco de Dados

Como mencionado anteriormente, o SGBD relacional utilizado será o PostgreSQL e o não-relacional será o MongoDB.

A base de dados utilizada será o SoccerWiki[4], a qual é constituída por jogadores de futebol, times e campeonatos.

4. Metodologia

Será feita uma pesquisa bibliográfica sobre as duas abordagens de SGBD. Em seguida, os dados serão coletados do site do SoccerWiki e os dois bancos de dados serão montados utilizando o PostgreSQL e o MongoDB (relacional e não-relacional, respectivamente).

Para a realização da análise entre o PostgreSQL e MongoDB, será utilizado o Apache JMeter[5], uma ferramenta open source, desenvolvida em Java[6], com o objetivo de medir desempenho. Dessa forma, serão feitas simulações de requisições aos SGBD, alterando, para cada um, o valor dos parâmetros referentes ao número de usuários que farão as requisições ao mesmo tempo, o número de repetições de requisições de cada usuário e o intervalo de tempo entre as requisições. Ao fim dos experimentos, os resultados obtidos no SGBD relacional serão comparados com os obtidos no SGBD não relacional.

Terminado o experimento, uma monografia será escrita, afim de registrar as observações coletadas.

5. Cronograma

Abaixo é apresentada uma proposta de cronograma para a realização do experimento e escrita do documento.

	Setembro		Outubro		Novembro		Dezembro	
Pesquisa e definição do escopo								
Pesquisa Bibliográfica								
Desenvolvimento do Ambiente/ Experimento								
Preparação do Documento								
Preparação da Apresentação Oral								
Apresentação Oral								

6. Proposta de Estrutura do Trabalho

Para este trabalho, a estrutura de documento proposta é a seguinte:

- 1. Introdução:** Neste capítulo, serão abordadas as motivações para a área de pesquisa de NOSQL, bem como uma visão geral sobre o assunto do trabalho e uma introdução do restante do documento.
- 2. Revisão Bibliográfica:** O uso de SGBD relacionais com quantidades muito grande de dados podem gerar problemas de eficiência no processamento, custo, escalabilidade, etc.[7]
A proposta de SGBD não-relacionais é solucionar a questão

da escalabilidade no armazenamento e processamento. O objetivo deste capítulo é trabalhar os conceitos de SGBD relacionais e não-relacionais.

3. Tecnologias, Dados e Plano de Experimento: Capítulo dedicado a introduzir as tecnologias utilizadas no experimento, como os SGBD PostgreSQL e MongoDB e a ferramenta JMeter, apresentar os dados utilizados e como eles foram modelados em cada SGBD e também, será detalhado o plano de testes que será executado no trabalho. Basicamente, serão alterados os valores de três parâmetros: O número de usuário fazendo requisições ao mesmo tempo, o número de vezes que cada usuário irá repetir as requisições e o intervalo entre duas requisições de um usuário. Serão escolhidos três valores para cada uma dessas variáveis e, para cada SGBD, serão utilizadas no experimento todas as combinações possíveis dos valores das variáveis, afim de obter 27 situações diferentes para cada SGBD. Espera-se, também, avaliar em quais situações é mais viável utilizar um SGBD não-relacional a um relacional.

4. Execução dos Experimentos e Análise de Resultados: Este capítulo aborda a execução dos experimentos descritos no capítulo anterior e analisa os resultados obtidos.

5. Conclusão: Aqui será apresentada a conclusão do trabalho, destacando as contribuições obtidas, citando as limitações e sugerindo trabalhos futuros.

Em seguida serão listadas as referências bibliográficas utilizadas no trabalho.

7. Referências

[1] DIANA, M.de; GEROSA. M.A. NOSQL na Web 2.0: Um Estudo Comparativo de Bancos Não-Relacionais para Armazenamento de Dados na Web 2.0. IX Workshop de Teses e Dissertações em Banco de Dados. São Paulo, 2010.

[2] PostgreSQL. Disponível em:<<http://www.postgresql.org>> Acesso em 2 de julho de 2013.

[3] MongoDB. Disponível em:<<http://www.mongodb.org>> Acesso em 2 julho de 2013.

[4] SoccerWiki. Disponpivel em <<http://pt-br.soccerwiki.org/wiki.php>> Acesso em 2 de julho de 2014.

[5] JMeter. Disponível em <<https://jmeter.apache.org>> Acesso em 2 de julho de 2014.

[6] Java. Disponível em <https://www.java.com/pt_BR> Acesso em 2 de julho de 2014.

[7] WANZELLER, Diogo Araújo Pacheco. **Investigando o Uso de Bancos de Dados Orientados a Documentos para Gerenciar Informações da Administração Pública**. 2013. 56 f. TCC (Graduação) - Curso de Licenciatura em Computação, Departamento de Departamento de Ciência da Computação, Universidade de Brasília, Brasília, 2013.

Fernando da Fonseca de Souza