**Universidade Federal De Pernambuco**

Graduação em Ciência da Computação

Centro de Informática

**Big Data e o impacto nos bancos relacionais: uma análise comparativa entre o uso do MySQL e do ElasticSearch no tratamento de Big Data em dados do Facebook.**

Proposta de Trabalho de Graduação

**Aluno**: Fernando Miguel Barros Lins (fmbl@cin.ufpe.br)

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

Recife, 09 de Dezembro de 2013

Sumário

1. Contexto . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
2. Motivação . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
3. Objetivo . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
4. Metodologia . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
5. Cronograma . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
6. Possíveis Avaliadores . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
7. Referências . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
8. Assinaturas . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

1. Contexto

Estamos vivenciando um momento onde o volume de dados trafegados na rede mundial de computadores é enorme. Isso graças a popularização das mídias sociais, engenhos de buscas, empresas criando e/ou migrando seus serviços para web. Alguns denominam esse fenômeno de web 2.0, e por causa do grande volume de dados, as plataformas tradicionais não estão conseguindo dar conta de tanta informação mantendo uma boa eficiência.

Com isso surgiu o que chamamos de Big Data, que tem como foco o grande armazenamento de dados e maior velocidade. Embaladas nesse novo conceito, empresas começaram a rever seus sistemas baseados em bancos de dados relacionais e a se questionarem sobre a capacidade dessas bases de acompanhar o processo evolutivo.

2. Motivação

A motivação surgiu através do projeto que trabalho no estágio. O projeto consiste em ser um ferramenta para monitorar e analisar o volume de buzz e criar ações de relacionamento com os consumidores no Facebook. Para isso, há crawlers que buscam os posts/comentários/likes (buzzes) das páginas do Facebook cadastradas. O projeto utiliza o MySQL e há um gargalo no tempo de resposta por causa do MySQL. Desta forma, começamos a procurar por uma nova ferramenta/sistema que substituísse o MySQL a fim de melhorar o desempenho de resposta (real-time).

3. Objetivo

Esse trabalho tem como objetivo realizar uma análise comparativa entre a utilização de banco de dados relacionais e novos sistemas de buscas de informações não relacionais no tratamento de grandes volumes de dados em tempo real com alto desempenho.

O sistema de banco de dados relacional utilizado na comparação será o MySQL e o não relacional será o ElasticSearch que é uma solução Schema Free, ou seja, sem esquema de dados definido.

Será usado com fonte de dados o Facebook, pois possui um imenso volume de dados disponível através de sua API.

4. Metodologia

O objetivo desse trabalho é uma analise comparativa, desta forma, para realizarmos a comparação entre o ElasticSearch e o MySQL será utilizado o Apache JMeter. Esta ferramenta é open source, feita em java e foi desenvolvida para, entre outras coisa, medir desempenho. Assim, serão feitos teste de latência, numero alto de acesso simulando vários usuário fazendo requisições.

5. Cronograma

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Dezembro | | | | Janeiro | | | | Fevereiro | | | | Março | | | |
| Coletar dados do Facebook |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Analisar desempenho |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Comparar desempenho |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Escrita da monografia |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Preparar apresentação |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Apresentação |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

6. Possíveis Avaliadores

* Ana Carolina Salgado
* Robson

7. Referências

VIEIRA, Marcos; DE FIGUEIREDO, Josiel; LIBERATTI, Gustavo; VIEBRANTZ, Alvaro. Disponível em: <http://data.ime.usp.br/sbbd2012/artigos/pdfs/sbbd\_min\_01.pdf>.

DUMBILL, Edd. **What is Big Data?.** O’Reilly Media. 2012

Mysql. Disponível em: <http://www.mysql.com/>.

Elasticsearch. Disponível em: <http://www.elasticsearch.org/>.

Apache JMeter. Disponível em: <https://jmeter.apache.org>.

Facebook. Disponível em: <http://www.facebook.com>.

Facebook API. Disponível em: <https://developers.facebook.com/docs/graph-api/>.

LOSCIO, Bernadette; OLIVEIRA, Hélio; PONTES, Jonas. Disponível em:

<http://www.addlabs.uff.br/sbsc\_site/SBSC2011\_NoSQL.pdf>.

DE DIANA, Mauricio; GEROSA, Marco. Disponível em:

<http://www.lbd.dcc.ufmg.br/colecoes/wtdbd/2010/sbbd\_wtd\_12.pdf>.

8. Assinaturas

Fernando da Fonseca de Souza

Orientador

Fernando Miguel Barros Lins

Aluno