



FUNDAÇÃO UNIVERSIDADE FEDERAL DO VALE DO SÃO FRANCISCO
COLEGIADO DE ENGENHARIA DE COMPUTAÇÃO

Avenida Antonio Carlos Magalhães, 510 – Santo Antônio CEP: 48902-300
Juazeiro/BA, Tel/Fax: (74) 2102-7636 - Juazeiro - BA
www.univasf.edu.br/~ccomp

Formulário de matrícula de TCC I

Título do projeto de pesquisa

Avaliação de banco de dados NoSQL para geração de séries temporais

Nome do discente

Jonas Henrique Alves Rodrigues

Nome do orientador e co-orientador

Rosalvo Ferreira de Oliveira Neto

Resumo e palavras-chave

NoSQL é um termo genérico para uma classe definida de banco de dados não-relacionais que rompe uma longa história de banco de dados relacionais com propriedades ACID. Tendências em arquiteturas de computadores, como a computação na nuvem e a necessidade crescente de prover serviços escaláveis, estão pressionando bancos de dados numa direção onde eles necessitam oferecer escalabilidade horizontal. Bancos de dados NoSQL armazenam os dados com técnicas que visam atender a esse requisito. Este estudo pretende comparar a performance entre banco de dados NoSQL e relacional na geração de séries temporais, uma série temporal é uma coleção de observações feitas sequencialmente ao longo do tempo. As séries temporais existem nas mais variadas áreas de aplicação, como: finanças, marketing, economia, ciências sociais, meteorologia, epidemiologia.

NoSQL, Big Data, Séries Temporais

Qualificação do problema a ser abordado

Big Data é um termo amplamente utilizado na atualidade para nomear conjuntos de dados muito grandes ou complexos que os aplicativos de processamento de dados tradicionais ainda não conseguem lidar como a geração de uma série temporal. Big Data possui três características: Volume, Variedade e Velocidade, também conhecida como 3Vs. Os desafios desta área incluem: análise, captura, curadoria de dados, pesquisa, compartilhamento, armazenamento, transferência, visualização e informações sobre privacidade dos dados. Sistemas de gerenciamento de banco de dados relacional, muitas vezes têm dificuldade em lidar com grandes volumes de dados, pois isto requer o trabalho de software paralelo, rodando em dezenas, ou até mesmo centenas de servidores.

Justificativa

Os bancos de dados NoSQL surgiram com a web 2.0, com crescente popularização das redes sociais, a geração de conteúdo por dispositivos móveis bem como o número cada vez maior de pessoas e dispositivos conectados, faz com que o trabalho de armazenamento de dados com o objetivo de utilizá-los em ferramentas analíticas, comece a esbarrar nas questões de escalabilidade e custos de manutenção desses dados. Bancos de dados relacionais escalam, mas quanto maior o tamanho, mais custoso se torna essa escalabilidade, seja pelo custo de novas máquinas, seja pelo aumento de especialistas nos bancos de dados utilizados, já os não relacionais permitem uma escalabilidade mais barata e menos trabalhosa, pois não exigem máquinas extremamente poderosas e sua facilidade de manutenção permite que um número menor de profissionais seja necessário.

Objetivo a ser alcançado

Este trabalho tem como objetivo responder a seguinte pergunta de pesquisa:

"Existe diferença de desempenho na geração de séries temporais entre banco de dados relacional e NoSQL?"

Metodologia a ser empregada

- Pesquisa dos bancos de dados NoSQL;
- Pesquisa de benchmark de séries temporais;
- Estudo do benchmark;
- Implementação banco de dados relacional;
- Implementação NoSQL;
- Análise dos dados;

Referencial bibliográfico

R. Cattell, "Scalable sql and nosql data stores," SIGMOD Rec., vol. 39, no. 4, pp. 12-27, 2011.

C.-W. Tsai, C.-F. Lai, H.-C. Chao, and A. Vasilakos, "Big data analytics: a survey," Journal of Big Data, vol. 2, no. 1, pp. 1-32, 2015.

D. Han and E. Stroulia, "A three-dimensional data model in HBase for large time-series dataset analysis," 2012 IEEE 6th International Workshop on the Maintenance and Evolution of Service-Oriented and Cloud-Based Systems (MESOCA), Trnto, 2012, pp. 47-56.

Y. Li and S. Manoharan, "A performance comparison of sql and nosql databases," in Communications, Computers and Signal Processing (PACRIM), 2013 IEEE Pacific Rim Conference on. IEEE, 2013, pp. 15-19.

Cronograma de atividades

	Pesquisa banco de dados NoSQL	Pesquisa Benchmark	Estudo do benchmark	Implementação banco de dados relacional	Implementação NoSQL	Análise Comparativa dos dados
Semana 1	x					
Semana 2	x					
Semana 3		x				
Semana 4		x				
Semana 5			x			
Semana 6				x		
Semana 7				x		
Semana 8					x	
Semana 9					x	
Semana 10					x	
Semana 11						x
Semana 12						x

Necessidades e disponibilidade de recursos e infraestrutura para o desenvolvimento deste projeto

Rosalvo Ferreira de Oliveira Neto
Orientador(a)

Jonas Henrique Alves Rodrigues
Aluno(a)

Juazeiro-BA, 08/12/2017.