

Dados estruturados em uma estrutura Event Sourcing: um estudo comparativo MongoDB x SQL Server

Matheus da Costa Oliveira

RESUMO

Este relatório apresenta (máximo de 20 linhas).

Apresente rapidamente o objetivo do trabalho, o desenvolvimento e os resultados obtidos.

Palavras-chave

MongoDB, SQL Server, Avaliação de performance

I. INTRODUÇÃO

1. Contexto

A escolha de um banco de dados é estrutural na criação de um software. Deve-se levar em conta, por exemplo, a compatibilidade do tipo de dados - relacional, documentos, objetos, etc – com os dados a serem armazenados, a escalabilidade, o custo e a capacidade de obtenção de mão de obra qualificada. Com a ampla adoção da computação em nuvem, escalabilidade e custo se tornaram ainda mais decisivos. Se por um lado a nuvem facilita o gerenciamento de infraestrutura, por outro o valor cobrado pelas empresas mantenedoras das ferramentas é bastante elevado, principalmente no Brasil, pois os preços são comumente cobrados em dólar.

A maior parte das empresas utilizam bancos de dados relacionais, tendo o SQL Server figurando entre os mais populares (STACK OVERFLOW, 2023). O SQL Server é suportado pela Microsoft usufruindo tanto do suporte da empresa quanto da comunidade. Isso, combinado com a disponibilidade de versões gratuitas para uso pessoal e educacional, facilitam a adoção por estudantes e garante uma parcela considerável de mão de obra qualificada.

Também amplamente adotado (STACK OVERFLOW, 2023), o MongoDB é um banco de dados orientado a documentos, e também possui suporte massivo da comunidade (STACK OVERFLOW, 2023). O MongoDB também possui versões gratuitas.

2. Objetivo

O objetivo deste trabalho consiste em observar e comparar a performance de leitura e escrita de dados gerados por uma aplicação *event sourcing* entre dois tipos de estrutura em um banco de dados, a Relacional, utilizando SQL Server, e a Orientado a Documentos, utilizando o MongoDB.

3. Motivação

Diversas empresas ou entidades que possuem sistemas legados, cujos dados estão armazenados em um banco de dados Relacional, ao criarem novos sistemas ou atualizar processos, optam por continuar no mesmo banco de dados devido a diminuição de custos e aproveitamento de conhecimento. Porém, a escolha adequada de um banco de dados garante à organização facilidade de implementação, manutenção, melhor utilização de recursos computacionais, escalabilidade e custo.

Os benefícios de uma mudança de tipo de banco de dados podem suplantam o custo de aprendizado e treinamento da equipe.

4. Materiais

Este trabalho foi realizado utilizando os seguintes recursos tecnológicos, modelos ou práticas:

- Sistema de Gerenciamento de Banco de Dados SQL Server 2022;
- IDE SQL Server Management Studio;
- Sistema de Gerenciamento de Banco de Dados MongoDB Versão 7;
- Docker Engine;
- Docker Hub;
- Plataforma de Desenvolvimento de Software .NET 8;
- Linguagem de Programação C#;
- IDE Visual Studio Code;

5. Metodologia do Trabalho

Este trabalho foi realizado nas seguintes etapas:

- Criação de um ambiente idêntico entre os dois bancos de dados, utilizando Docker;
- Criação de um programa gerador de transações aleatórias, de inserção e consulta, a fim de simular um ambiente de produção real;

- Execução do programa com diferentes escalas de transações e armazenamento dos resultados;
- Análise dos resultados comparando a performance dos dois bancos de dados.

II. REVISÃO DE TECNOLOGIAS E PRÁTICAS

1. SQL Server e bancos relacionais

Bancos de dados relacionais utilizam a linguagem de consulta SQL. Armazenam os dados em tabelas, que consistem em um conjunto de registros, sendo cada registro composto por campos, também conhecidos por colunas (ELMASRI et al, 2005). Essa estrutura tabular entende que os dados ali armazenados possuem o mesmo tipo e o mesmo número de campos, sendo esses dados, portanto, estruturados.

Com o objetivo de diminuir ou eliminar a repetição de dados, bancos relacionais são normalmente normalizados, o que acarreta a criação de múltiplas tabelas, com campos que apontam para um outro registro em outra tabela. Esses relacionamentos entre tabelas obrigam o uso de joins na consulta dos dados. Quanto maior o número de tabelas presentes na consulta e maior o número de registros em cada uma, mais complexo e custoso ela é.

O SQL Server é um dos mais populares bancos relacionais do mercado, com ampla aceitação entre profissionais e estudantes (STACK OVERFLOW, 2024). Possui diversos trabalhos que provam sua eficiência em consultas, principalmente envolvendo o uso de *joins* (PARKER, 2013).

2. NoSQL e MongoDB

Diferente dos bancos relacionais, os bancos NoSQL não obrigam que os dados armazenados sejam estruturados, não possuem tabelas nem relacionamentos. São altamente escaláveis e tolerantes à falhas, trabalhando com eficiência com grandes volumes de dados (PARKER, 2013). Existem muitos tipos de bancos NoSQL, mas neste trabalho utilizaremos o MongoDB.

O MongoDB é um banco de código aberto orientado a documentos, armazenando dados em formato JSON. Cada documento contém todas as informações relacionadas a ele, retirando assim a necessidade de joins, mas aumentando o uso de armazenamento.

3. Event Sourcing

É comum na tecnologia da informação a necessidade de geração de dados de auditoria ou *logs*. Esses dados garantem uma rastreabilidade, uma vez que é possível analisar todos os eventos que levaram ao estado atual dos dados.

No desenvolvimento de software, um padrão possível de ser utilizado é o *Event Sourcing*. Esse padrão define a captura de todos os eventos de um determinado sistema ou entidade como uma sequência imutável (FOWLER, 2023). Assim, a rastreabilidade é garantida, já que não é possível a alteração dos dados já inseridos.

III. DESENVOLVIMENTO

Nesta seção são apresentadas detalhadamente as etapas de desenvolvimento do trabalho e etc
Crie quantas seções forem necessárias

1. Abc

2. Def

Crie quantas seções forem necessárias

IV. RESULTADOS E ANÁLISE

Nesta seção são apresentados detalhadamente os resultados do trabalho

Crie quantas seções forem necessárias

1. aaaaa

2. aaaaa

V. CONCLUSÃO

O trabalho realizado permite avaliar e que

1. Dificuldade Encontradas

Nesta seção são apresentadas as dificuldades encontradas para a realização deste trabalho. SEÇÃO OBRIGATÓRIA

2. Aplicabilidade do Trabalho

Nesta seção são apresentados contextos em que este trabalho pode ser aplicado. Estes contextos podem ser empresas, setores, áreas ou tipos particulares de problemas e situações SEÇÃO OBRIGATÓRIA

VI. BIBLIOGRAFIA

Coloque as indicações bibliográficas seguindo do padrão ISO

ELMASRI, Ramez e Shamkant NAVATHE. Sistemas de Banco de Dados. 4a ed. Addison Wesley, 2005. ISBN 9788588639171.

FOWLER, Martin. Event Sourcing. martinowler.com [em linha]. [sem data] [consult. 20 mar 2024]. Disponível em: <https://martinfowler.com/eaDev/EventSourcing.html>

PARKER, Zachary, Scott POE e Susan V. VRBSKY. Comparing NoSQL MongoDB to an SQL DB. In: the 51st ACM Southeast Conference [em linha]. New York, New York, USA: ACM Press, 2013 [consult. 21 mar 2024]. ISBN 9781450319010. Disponível em: doi:10.1145/2498328.2500047

STACK OVERFLOW. Stack Overflow Developer Survey 2023. Stack Overflow [em linha]. [sem data] [consult. 21 mar 2024]. Disponível em: <https://survey.stackoverflow.co/2023/#most-popular-technologies-database-prof>