



Estudo de bancos de dados NoSQL como solução alternativa ao modelo relacional.

Gisele Busichia Baioco (PQ), Lucas Watanuki (IC).

Resumo

Este trabalho consistiu na análise técnica e comparativa dos principais Sistemas de Gerenciamento de Bancos de Dados (SGBD) não relacionais, os quais possuem características como alta escalabilidade horizontal, *schemas* flexíveis e variados tipos de arquitetura de acordo com a categoria do banco de dados: (i) Documentos, (ii) Chave-Valor, (iii) Colunas e (iv) Grafos. Por meio dessa análise foi possível mapear os SGBD não relacionais dentro de cada categoria e obter um fluxograma que pode sistematizar a escolha de um determinado SGBD NoSQL (Not Only SQL).

Palavras Chave: escalabilidade, SGBD, NoSQL.

Introdução

Segundo Fowler (2015), o NoSQL é um tipo de banco de dados que não requer um schema rigoroso para manipular registros e seus metadados em si, não utiliza o modelo matemático dos bancos de dados relacionais, e que pode ser utilizado de forma mais distribuída em um hardware comum. De modo geral, grande parte desses bancos são open source (código aberto), altamente escaláveis, possuem *schemas* flexíveis e API (Application Programming Interface) simples (VYAWAHARE; KARDE; THAKARE, 2017). O ponto a se destacar é que existem diversas categorias de bancos não relacionais, e dentro de cada categoria temos vários SGBD com características semelhantes, mas que diferem entre si em quesitos de configuração, APIs, usabilidade, interfaces de gerenciamento, desempenho em consultas e operações de escrita e leitura de dados e segurança de acesso a metadados e schemas.

Portanto, com o intuito de otimizar o processo de decisão ao especificar um modelo de banco de dados não relacional e seu respectivo SGBD para um determinado projeto ou aplicação, foi proposta uma avaliação sobre cada categoria de bancos NoSQL de acordo com os seguintes critérios: (i) Interface Gráfica para gerenciamento dos dados, (ii) Operações CRUD (*Create, Read, Update, Delete*), (iii) Ambientes Suportados, (iv) Segurança e (v) Aplicabilidade.

Resultados e Discussão

Com base nas análises dos SGBDs dentro de cada categoria de banco de dados NoSQL, foi possível mapear de maneira sistemática as características particulares de cada arquitetura e estrutura de dados, observado na Figura 1, tudo de acordo com os critérios estabelecidos anteriormente. Com as informações coletadas e a partir dos testes realizados em cada SGBD, foi possível sintetizar na Tabela 1 as propriedades e atributos avaliados, bem como o desenvolvimento de um fluxograma, Figura 2, que contribui de maneira mais sistemática a escolha de um determinado SGBD NoSQL dentro do contexto de aplicações modernas que exigem flexibilidade, escalabilidade e alta disponibilidade de dados.



Figura 1. Fluxograma para apoio na escolha de SGBD não relacional a partir de suas características

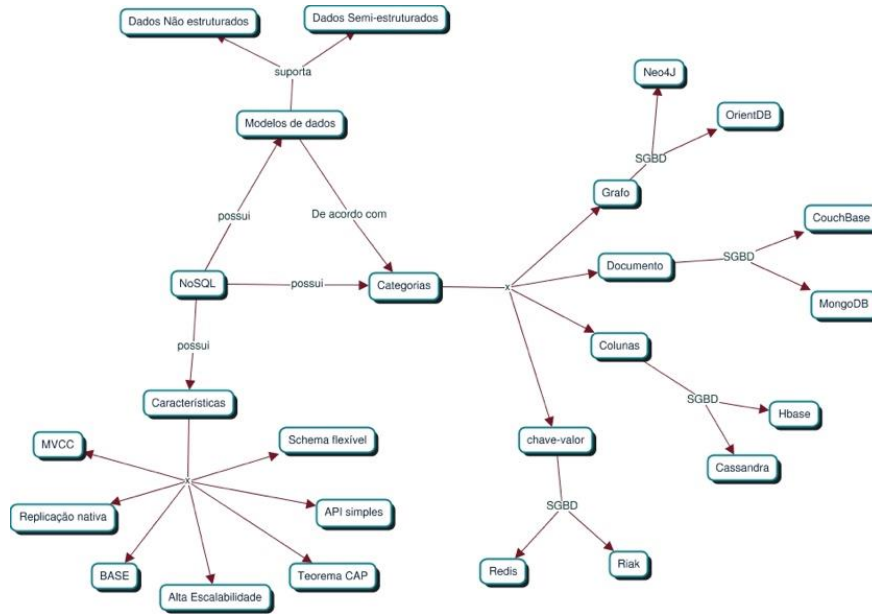


Figura 2. Fluxograma para apoio na escolha de SGBD não relacional a partir de suas características

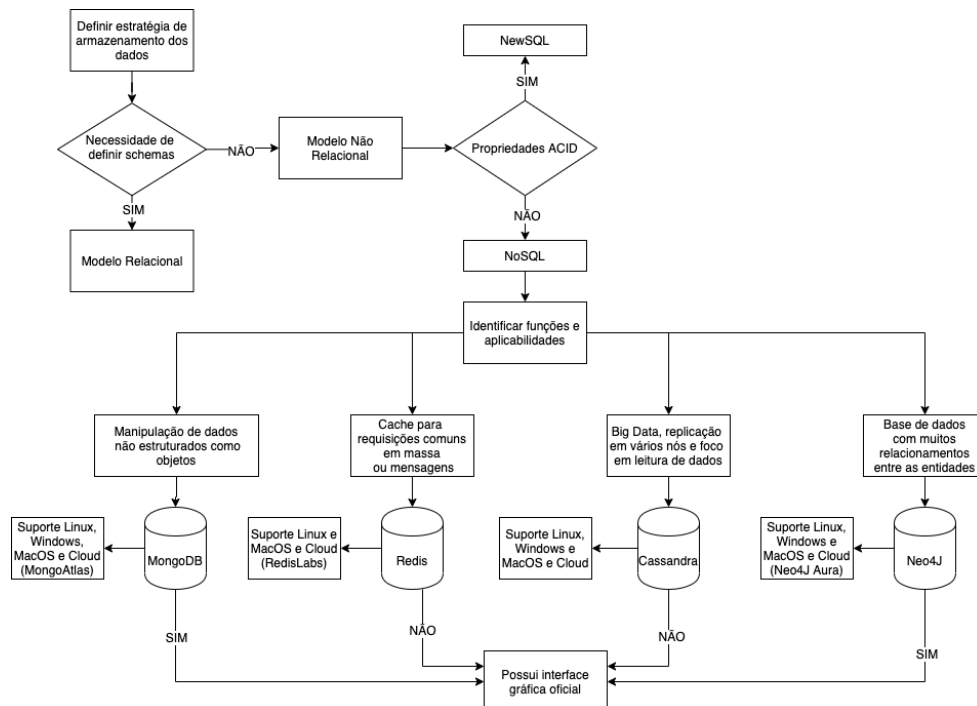




Tabela 1 - Síntese das características de SGBD NoSQL

SGBD	Interface	Ambiente	Linguagem de acesso	Segurança	Aplicabilidade
MongoDB (Documentos)	MongoDB Compass	Linux, MacOS, Windows e Cloud.	Métodos em Javascript pelo MongoShell ou adaptado a linguagem de programação cliente	<i>Role-Based Access Control (RBAC)</i> para controle de acesso aos <i>schemas</i> , documentos e operações CRUD	Pode ser aproveitado em aplicações que manipulam diversos dados semi-estruturados ou não estruturados.
Redis (Chave-Valor)	Possui interface de terceiros, porém pago. Exemplo: Redis Desktop Manager	Linux, MacOS e Cloud.	Podemos realizar operações CRUD pelo redis-cli através dos seguintes comandos: SET, GET, DEL	Redis ACL (Access Control List) – permite limitar conexões em termos de comandos que podem ser executados e as chaves que podem ser visualizadas.	Ideal para usar como cache de requisições ou intermediário de mensagens e fila.
Cassandra (Colunas)	Não possui	Linux, Windows e MacOS.	<i>Cassandra Query Language (CQL)</i> semelhante ao SQL, porém não possui cláusulas JOIN, GROUP BY ou FOREIGN KEY	<i>Role-Based Access Control (RBAC)</i> para acesso a <i>Keyspaces</i> , tabelas e funções.	É uma ótima opção utilizar o Cassandra quando há exigência na replicação de dados em massa, onde há uma sobrecarga sobre o banco de dados e maior tendência a leituras em detrimento de escritas.
Neo4J (Grafos)	Neo4J Desktop e Neo4J Browser	Linux, Windows, MacOS e Cloud.	Cypher	Schema-based security - Oferece segurança a nível relacionamentos e nós em grafos (schemas)	Base de dados que contenham muitos relacionamentos e seja necessário realizar uma consulta complexa havendo a necessidade de múltiplas junções entre as diversas tabelas

Também foram testados alguns SGBDs NewSQL, que possuem características tanto do modelo relacional quanto relacional, ou seja, oferecem desempenho escalável para cargas de trabalho de leitura e escrita ao Processamento de Transações On-Line (OLTP), enquanto mantém as propriedades ACID, isto é, garantias de Atomicidade, Consistência, Isolamento e Durabilidade de um SGBD Relacional (CÁ, 2018). Os SGBDs NewSQL avaliados e testados foram: NuoDB, CockroachDB e o VoltoDB.



Conclusões

Considerando as atuais demandas de projetos dentro do contexto de computação em nuvem, o aumento exponencial do volume de dados que são transitados através da internet, bem como as inúmeras aplicações que precisam manipular dados não estruturados e semi-estruturados, este projeto procurou focar em avaliar qualitativamente os principais SGBD de cada categoria dos bancos não relacionais de acordo com critérios bem definidos e que são amplamente discutidos ao escolher determinado SGBD, abordando aspectos sobre interface gráfica, ambiente de execução, linguagem de manipulação dos dados, aspectos de segurança e aplicabilidade.

As características dos SGBD avaliados foram evidenciadas através da análise das suas respectivas documentações e principalmente a partir de testes em máquina local e contêineres, devido a algumas incompatibilidades de sistema operacional.

A partir da análise descrita anteriormente, foi possível sintetizar as informações obtidas e assim elaborar um diagrama que nos permite ter uma visão macro acerca dos principais SGBD NoSQL e NewSQL a fim de apoiar na escolha e especificação dos mesmos para uma determinada aplicação.

Sendo assim, considerando todos os mapeamentos, diagramas e resultados obtidos, o projeto cumpre com seu objetivo de abordar analiticamente os principais SGBD alternativos ao modelo relacional, bem como proporcionar subsídios que auxiliam na escolha de um modelo de banco de dados no contexto de aplicações modernas, e que muitas vezes são consideradas políglotas, isto é, que exigem flexibilidade em relação a modelagem, escalabilidade e disponibilidade de dados em diferentes camadas.

Agradecimentos

Agradeço a Universidade Estadual de Campinas, aos professores, aos departamentos de orientação, direção e administração da Faculdade de Tecnologia por fornecerem um sólido ambiente de aprendizado e desenvolvimento.

A minha professora orientadora Gisele Busichia Baioco, pela atenção, empenho e dedicação ao projeto de pesquisa.

E por fim, ao PIBIC - Programa Institucional de Bolsas de Iniciação Científica e ao CNPq, pela confiança depositada nesse projeto, permitindo a expansão e aprimoramento da minha formação acadêmica.

Referências

FOWLER, A. NoSQL For Dummies. 1. ed. [s.l.] John Wiley & Sons, 2015.

VYAWAHARE, H.; KARDE, D. P.; THAKARE, D. V. M. NoSql Database. International Journal of Evolutionary Scientific Research and Technology, v. 1, n. 1, p. 11–14, 2017.

CÁ, J, L;. Uma Abordagem Para Modelagem De Desempenho Para Sistemas De Banco De Dados Newsq E Comparação De Modelos Gerados Com Diferentes Técnicas De Aprendizado De Máquina. Universidade Estadual Do Ceará, 2018.