

# **BANCOS DE DADOS NOSQL COMO SUPORTE A ANÁLISE DE DADOS NÃO ESTRUTURADOS**

Guilherme de Cleve FARTO

*guilherme.farto@gmail.com*

Gabriel Matheus Bernardo da SILVA

*gabrielbernardo14@outlook.com*

**RESUMO:** Com a crescente dos dados e a necessidade de cada vez mais buscar ferramentas que facilitam o gerenciamento e trouxessem melhor desempenho, assim surgiu o conceito de NoSQL. Tendo como base o uso de dados não estruturados, torna mais prático armazenar e manusear dados que possuem modelos diferentes, sendo mais vantajosos para cada tipo de uso. Este, por sua vez, armazena um objeto em formato JSON e traz consigo uma estrutura não obrigatória, assim variável e flexível. Suas principais vantagens podem ser definidas como escalabilidade, flexibilidade e melhor desenvolvimento com grande número de dados.

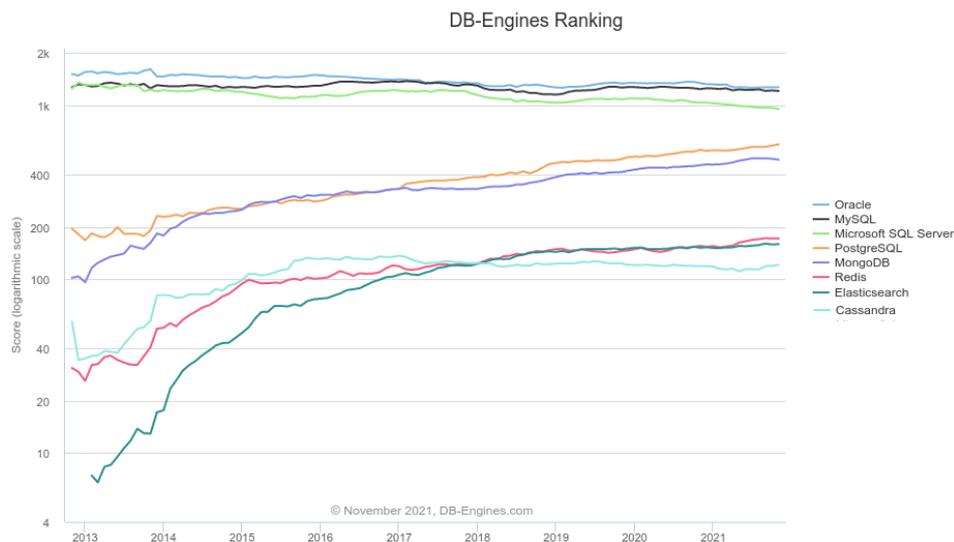
**PALAVRAS-CHAVE:** NoSQL; Documentos; Open Source; JSON; escalabilidade;

**ABSTRACT:** With the growth of data and the need to increasingly seek tools that facilitate management and bring better performance, that's how the concept of NoSQL emerged. Based on the use of unstructured data, it makes it more practical to store and handle data that have different models, being more advantageous for each type of use. This, in turn, stores an object in JSON format and brings with it a non-mandatory structure, thus variable and flexible. Its main categories can be defined as scalability, flexibility and better development with large amount of data.

**KEYWORDS:** NoSQL; Documents; Open Source; BSON; scalability;

## 1. Introdução

Como se pode perceber, partindo de uma observação no site DB-ENGINES, o crescimento dos bancos de dados NoSQL é constante e a cada ano que passa, sua projeção de crescimento é maior, como na Figura 1.



**Figura 1:** Ranking de crescimento de bancos de dados (In: DB-ENGINES, nov 2021)

Como apresentado na figura, é possível verificar os principais bancos de dados relacionais (Oracle, MySQL, SQLServer e PostgreSQL) e os principais NoSQL (MongoDB – modelo documentos, Redis – modelo chave-valor, ElasticSearch – modelo de busca e Cassandra – modelo de colunas). Os bancos de dados NoSQL possuem diversos modelos diferentes de dados, onde cada ferramenta usa algum em específico que é destinado para uma forma de manipulação, onde cada um dos principais será tratado na seção a seguir.

O conceito de NoSQL trouxe ao mercado uma forma que promete alcançar um grande número de aplicações, com qualidade, e com soluções mais adequadas ou complementando as já existentes dos bancos relacionais (BOGONI, 2010).

Os gerenciadores de NoSQL estão se tornando grandes opções principalmente quando se trata de *Big Data*, onde seu desempenho está sendo superior aos relacionais além de possuírem mecanismos de particionamento e replicação de dados, com configurações modernas e permitirem a adoção de diferentes níveis de consistência entre servidores, retirando diversos obstáculos (RODRIGUES, 2019).

## 2. NoSQL

O termo NoSQL se refere a tipos de bancos de dados não relacionais que armazenam dados por meio de diversos formatos, de acordo com a ferramenta utilizada para o gerenciamento. Estes podem ser manipulados por meio de APIs de diversas linguagens, e assim, se surge o sentido da palavra, que é “não apenas SQL” (ORACLE, 2021).

Este tipo de banco de dados apresentam um esquema dinâmico, com alto desempenho em consultas e de valor gratuito, além de ser altamente flexível, com grande escalabilidade e tem suporte por uma comunidade totalmente independente e de fácil acesso a informação (VICTÓRIA, 2019).

Os tipos de modelos de bancos de dados NoSQL mais comuns e suas principais características são:

- Chave-valor: se utiliza por meio de uma tabela de hash que fazem o pareamento de chaves e valores de dados, estes são altamente particionáveis e permitem grande escalabilidade horizontal com bastante performance;
- Documento: se armazena os dados em documentos em formato JSON onde estes pertencem a coleções, onde são compatíveis com o tipo chave-valor em um único documento e pode se consultar qualquer atributo que fizer parte do documento;
- Gráfico: se tem como base o conceito de nós e bordas para representação dos dados que estão conectados, onde são bastante usados em redes sociais e em detecção de fraudes;
- Colunar: Se realiza consultas por meio de linhas de forma aleatória, podendo consultar colunas específicas.

Esses modelos são variados de acordo com o gerenciador de banco escolhido para se trabalhar (AZURE, 2021).

No projeto realizado, o gerenciador de banco de dados NoSQL utilizada foi o MongoDB.

## 3. Open Source

Significa “Código Aberto”, ou seja, se refere ao código-fonte de uma aplicação, onde, em grande parte das vezes, se tem o compartilhamento do mesmo por meio de profissionais da área e possibilita que qualquer pessoa possa fazer modificações e melhorias no código para seu uso.

O crescimento no mercado desse uso teve seu início a cerca de 20 anos atrás e seu objetivo principal sempre foi buscar melhores performances do código onde, pessoas que estudam alguns pontos específicos de certas tecnologias pudessem modificar e implementar no código estas melhorias, tornando a aplicação cada dia melhor.

Nos dias atuais, há uma grande parte de aplicações com esse tipo de código, onde uma delas, bem conhecida, é o sistema operacional Linux com suas diversas versões. Outra aplicação bastante famosa é o GitHub, bem como redes sociais e outras aplicações espalhadas pelo mundo virtual (TOTVS, 2021).

## **4. Proposta de Desenvolvimento do Trabalho**

### **4.1. Objetivos**

Esta pesquisa de iniciação científica tem como objetivo principal estudar e aprofundar nos conhecimentos de bancos de dados NoSQL abordando comparações com bancos SQL e verificar seu funcionamento em uma aplicação.

### **4.2. Tecnologias e Recursos**

#### **4.2.1 MongoDB**

Atualmente é o gerenciador de banco de dados não relacional mais utilizado. O MongoDB foi criado em 2007, onde, segundo os fundadores, foi construído com base em experiências que os mesmos tinham com grande escala, sistemas robustos e alta disponibilidade, e assim criaram este novo tipo de banco de dados.

O MongoDB tem como principais características ser Open Source (ponto de grande vantagem dos bancos não-relacionais), tem alta performance, esquema aberto, além de possuir um ambiente para se trabalhar em nuvem (Atlas) e ser orientado a documentos, esses últimos dois tópicos serão apresentados a seguir (SILVA, 2021).

Atualmente está em sua versão 5.0.

##### **4.2.1.1 Atlas**

Trata-se de uma plataforma de dados de aplicativos multi-nuvem, onde se possui um conjunto integrado de bancos de dados em nuvem permitindo oferecer serviços de dados rápidos e simplificados por meio da nuvem, ou seja, sem se necessitar instalação de diversos softwares.

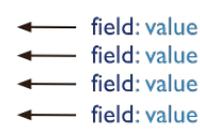
Possui alta facilidade e flexibilidade de uso para criação e análise de dados, com uma interface moderna e apresentação de gráficos, além do uso de *Data Lake* (mecanismo de consulta escalonável e sem servidor), além de possuir alto nível de segurança e alcance global (ATLAS, 2021).

#### 4.2.1.2 Orientação a Documentos

Este tipo de orientação trata cada documento como sendo um conjunto de uma ou mais propriedades, ou seja, um objeto, onde o MongoDB permite que dados relacionados possam ser incorporados em um único documento, assim, se pode trabalhar com documentos incorporados, que é quando se tem documentos embutidos em outro documento e realiza a manipulação destes (MANUAL MONGODB, 2021).

Cada dado presente em um documento tem sua representação um campo e um valor daquele campo, utilizando uma estrutura JSON, onde, em especial, o MongoDB faz uso do BSON. Uma representação de documento pode ser observada na Figura 2.

```
{
  name: "sue",
  age: 26,
  status: "A",
  groups: [ "news", "sports" ]
}
```



The diagram shows a JSON document with four fields: 'name', 'age', 'status', and 'groups'. Each field is followed by its value. To the right of each value, there is a blue arrow pointing left towards the field name, with the text 'field: value' next to it. This illustrates the mapping between a field and its corresponding value in a document.

**Figura 2:** Representação de Documento (In: MANUAL MONGODB, 2021)

#### 4.2.1.3 BSON

Se trata de uma codificação binária de documentos JSON, onde se tem a representação de dados que não fazem parte do JSON tradicional, como por exemplo os tipos Date e BinData.

Ele pode ser mais flexível que o JSON porém também pode ter uma eficiência de espaço menor, e tem como características: ser leve, eficiente e atravessável (BSON, 2021).

#### 4.2.2 Flutter

O Flutter é uma estrutura pertencente a empresa Google que tem código aberto, utilizado para a construção de aplicativos multiplataformas, compilados a partir de uma base única de código. Suas principais características são:

- Rápido: Isso se deve ao fato de o código ser compilado para máquinas ARM ou Intel utilizando JavaScript;

- **Produtivo:** Permite fazer uso de Hot Reload (é um recurso que injeta códigos-fonte na DVM (*Data Virtual Machine*)) e por meio da atualização do código a aplicação modifica instantaneamente;
- **Flexível:** É possível personalizar toda a aparência da aplicação, deixando-a da forma que se desejar.

Se pode utilizar para o desenvolvimento de aplicações para: Android, iOS, Web, Windows, macOS e Linux. Suporta diversas IDEs como VSCode e Android Studio, e faz uso da linguagem Dart para seu desenvolvimento (FLUTTER, 2021).

#### 4.2.2.1 Dart

É uma linguagem otimizada para se utilizar no desenvolvimento de aplicativos para diversas plataformas. Traz como objetivo oferecer uma linguagem de programação rápida e produtiva, além de ser flexível para ser usada com *frameworks* de diversos aplicativos (DART, 2021).

## 6.2. Testes e Resultados

Fazendo uso dos conceitos e ferramentas citados, foi possível desenvolver um estudo a respeito dos temas utilizados, além de ser possível realizar um curso a respeito do MongoDB, assim, conhecendo melhor esta ferramenta. Com isso, pode-se observar o comportamento desta ferramenta durante o manuseio de dados, onde se teve ambientes com pequena quantidades de dados, ambientes com dados de valor mediano, cerca de 50 mil e ambientes com grande volume de dados, cerca de 100 milhões.

Durante este estudo, se pode observar a grande flexibilidade e ganho de performance com o uso de bancos de dados não relacional, onde, após, foi inicializado, como teste, o desenvolvimento de uma pequena aplicação com o ambiente VSCode fazendo uso do Kit de Desenvolvimento Flutter. Pode-se perceber que o manuseio na linguagem Dart com o MongoDB foi mais fácil que com o MySQL, que é um banco relacional, além do desempenho do primeiro ser bem superior trabalhando com bases de dados semelhantes, onde, enquanto com o Mongo demorou cerca de 10s para processar a requisição, como MySQL foi cerca de 3x maior.

Em um aprofundamento maior, com o uso de indexes, os bancos não relacionais tornaram a experiência com os dados ainda mais rápida, e por meio de ferramentas como o MongoDB Compass, que é um ambiente gráfico da Empresa MongoDB, se pode ter uma visualização bem explícita de como os dados são manipulados, em seus scripts, principalmente de busca, além de se poder ver a quantidade de armazenamento

que está sendo ocupada por cada coleção juntamente com a quantidade de dados na mesma.

### 6.3. Trabalhos Futuros

Para trabalhos futuros podem ser sugeridos diversas modificações em aplicações que fazem uso, na prática, do manuseio com bancos de dados relacionais para que seja feita a migração para bancos não relacionais, podendo assim, com grande certeza, melhorar bastante o desempenho destas.

## 7. Conclusões

A tendência de crescimento dos bancos não relacionais é crescer, principalmente devido a integração com ambientes *Cloud*, onde, no projeto citado, existe o Atlas da empresa MongoDB.

Diante do estudo realizado, se pode concluir que os bancos de dados não relacionais vieram para complementar os bancos relacionais, onde cada um tem seus pontos vantajosos e de desvantagem.

Devido aos bancos não relacionais serem bastante recentes, ainda não possuem grande domínio e conhecimento no mercado, porém diversas grandes aplicações utilizadas mundialmente já fazem uso de bancos NoSQL, como: Facebook, Twitter, Digg e SourceForget. Com isso, se pode perceber que essa tecnologia tem grande potencial, principalmente quando se trata de alto volume de dados, fazendo assim, com que uma aplicação fique mais atrativa para uso devido a sua rapidez e eficiência com dados (DIANA; GEROSA, 2010).

## 8. Referências

ATLAS, MongoDB. **The multi-cloud application data platform**. MongoDB. Disponível em <<https://www.mongodb.com/pt-br/atlas>>. Acesso em 23 nov 2021.

AZURE. **Banco de Dados NoSQL – O que é NoSQL?**. Microsoft Azure. Disponível em <<https://azure.microsoft.com/pt-br/overview/nosql-database/>>. Acesso em 22 nov 2021.

BOGONI, Leandro P. **Por dentro do movimento NoSQL – SQL Magazine 83**. DevMedia. Disponível em <<https://www.devmedia.com.br/por-dentro-do-movimento-nosql-sql-magazine-83/18942>>. Acesso em 23 nov 2021.

BSON. **BSON**. BSON. Disponível em <<https://bsonspec.org/>>. Acesso em 21 nov 2021.

DART. **Dart overview**. Dart. Disponível em <<https://dart.dev/overview>>. Acesso em 21 nov 2021.

DB-ENGINES. **DB-Engines Ranking**. DB-ENGINES. Disponível em <<https://db-engines.com/en/ranking>>. Acesso em 24 nov 2021.

DIANA, Maurício de.; GEROSA, Marco Aurélio. **NoSQL na Web 2.0: Um Estudo Comparativo de Bancos Não-Relacionais para Armazenamento de Dados na Web 2.0**. In: Workshop de Teses e Dissertações em Banco de Dados, 9, 2010, Belo Horizonte.

FLUTTER. **Build apps for any screen**. Flutter. Disponível em <<https://flutter.dev/>>. Acesso em 20 nov 2021.

MANUAL MONGODB. **Introduction to MongoDB**. MongoDB. Disponível em <<https://docs.mongodb.com/manual/introduction/>>. Acesso em 23 nov 2021.

ORACLE. **O que é NoSQL?**. ORACLE. Disponível em <<https://www.oracle.com/br/database/nosql/what-is-nosql/>>. Acesso em 23 nov 2021.

RODRIGUES, Matheus Dornelas. **Avaliação de SGBDs NoSQL em ambientes de nuvem privada: uma abordagem baseada em modelagem estocástica**. 2019. 90p. Dissertação (Mestrado) – Centro de Informática – Universidade Federal de Pernambuco, Pernambuco, Recife, 2019.

SILVA, Fábio. **Como surgiu o MongoDB**. FabioSilva. Disponível em <<https://fabiosilva.com.br/2014/08/07/como-surgiu-o-mongodb/>>. Acesso em 22 nov 2021.

TOTVS. **Open Source (Código Aberto): Como funciona e suas vantagens**. TOTVS. Disponível em <<https://www.totvs.com/blog/developers/como-funciona-um-software-open-source/>>. Acesso em 24 nov 2021.

VICTÓRIA, Penélope. **Banco de dados NoSQL: um manual prático e didático**. GeekHunter. Disponível em <<https://blog.geekhunter.com.br/banco-de-dados-nosql-um-manual-pratico-e-didatico/>>. Acesso em 22 nov 2021.