

Escolha da base de dados

Nos últimos 20 anos, esforços foram realizados para armazenar eficazmente os dados de saúde em meios eletrônicos e disponibilizá-los a quem deles necessita. Estes dados, que incluem dados não estruturados, semiestruturados e estruturados, contêm todas as características de *Big Data*, como a quantidade, a diversidade, a rapidez e a autenticidade. Em circunstâncias típicas, são obtidos por utilizadores a vários níveis (utentes, profissionais de saúde, gestão hospitalar e governo), que realizam pesquisas na base de dados. Algumas destas requerem uma execução rápida para fazer face a emergências, tornando imperativa uma resposta rápida ^[1].

Normalmente, os bancos de dados relacionais são utilizados para guardar dados de saúde, mas não são capazes de gerir a vasta e variada diversidade deles. Neste cenário, várias abordagens NoSQL têm sido analisadas para administrar *Big Data* em diferentes aplicações, uma vez que o desempenho de consultas e a escalabilidade horizontal são aprimorados através do uso de bancos de dados NoSQL ^[1,2].

Num estudo realizado por Goli-Malekabadi et al., foi proposto um modelo que utiliza bancos de dados NoSQL para guardar informações de saúde. As soluções NoSQL baseadas em documentos foram escolhidas conforme o tipo de dados de saúde. O modelo em questão foi aplicado a um contexto de *cloud* para aproveitar as características de distribuição. As informações foram guardadas no banco de dados por meio de uma configuração chamada "sharding", que permite que o banco de dados seja um sistema distribuído. O modelo foi avaliado relativamente a um modelo relacional de dados, levando em conta fatores como o tempo de consulta, a preparação dos dados, a flexibilidade e a capacidade de expansão. Ademais, ficou comprovado que o desempenho do MongoDB superou o do servidor SQL convencional em aspetos como flexibilidade, preparação de dados e extensibilidade ^[2, 3].

Assim, dentre as várias opções de armazenamento NoSQL normalmente empregadas por desenvolvedores, o banco de dados MongoDB tem se destacado como uma das soluções mais empregadas ^[1].

Over the last 20 years, efforts have been made to effectively store health data in electronic media and make it available to those who need it. This data, which includes unstructured, semi-structured and structured data, contains all the characteristics of *Big Data*, such as quantity, diversity, speed and authenticity. In typical circumstances, they are obtained by users at various levels (patients, health professionals, hospital management and government), who carry out searches on the database. Some of these require rapid execution to deal with emergencies, making a quick response imperative ^[1].

Normally, relational databases are used to store health data, but they are not capable of managing the vast and varied diversity of these. In this scenario, various NoSQL approaches have been analyzed to manage Big Data in different applications, since query performance and horizontal scalability are improved using NoSQL databases ^[1,2].

In a study by Goli-Malekabadi et al, a model was proposed that uses NoSQL databases to store health information. Document-based NoSQL solutions were chosen according to the type of health data. The model in question was applied to a *cloud* context to take advantage of its distribution characteristics. The information was stored in the database using a configuration called “sharding”, which allows the database to be a distributed system. The model was evaluated in relation to a relational data model, considering factors such as query time, data preparation, flexibility and expandability. In addition, MongoDB outperformed the conventional SQL server in aspects such as flexibility, data preparation and extensibility ^[2,3].

Thus, among the various NoSQL storage options commonly employed by developers, the MongoDB database has stood out as one of the most widely used solutions ^[1].

Referências bibliográficas

- [1] Sen, P. S., & Mukherjee, N. (2023). An ontology-based approach to designing a NoSQL database for semi-structured and unstructured health data. *Cluster Computing*, 27(1), 1–18. <https://doi.org/10.1007/s10586-023-03995-y>
- [2] Celesti, A., Lay-Ekuakille, A., Wan, J., Fazio, M., Celesti, F., Romano, A., Bramanti, P., & Villari, M. (2020). Information management in IoT cloud-based tele-rehabilitation as a service for smart cities: Comparison of NoSQL approaches. *Measurement: Journal of the International Measurement Confederation*, 151(107218), 107218. <https://doi.org/10.1016/j.measurement.2019.107218>
- [3] Goli-Malekabadi, Z., Sargolzaei-Javan, M., & Akbari, M. K. (2016). An effective model for store and retrieve big health data in cloud computing. *Computer Methods and Programs in Biomedicine*, 132, 75–82. <https://doi.org/10.1016/j.cmpb.2016.04.016>