

# Adoção de SGBDs NoSQL em Empresas Brasileiras: um Levantamento Preliminar\*

Nicolas Atmatzides<sup>1</sup>, Marcos Bedo<sup>2</sup>, Daniel de Oliveira<sup>1</sup>

<sup>1</sup> Instituto de Computação – Universidade Federal Fluminense (IC/UFF)

<sup>2</sup>Instituto do Noroeste Fluminense – Universidade Federal Fluminense (INFES/UFF)

{nicolas\_atmatzides,marcosbedo}@id.uff.br, danielcmo@ic.uff.br

**Resumo.** Nos últimos anos foi observado um aumento na popularidade de Sistemas de Gerência de Banco de Dados (SGBD) NoSQL. Esses SGBDs comumente oferecem flexibilidade na representação dos dados e escalabilidade horizontal facilitada. Apesar de ser de conhecimento comum que grandes empresas internacionais adotam essas soluções, não é claro o cenário de adoção de SGBDs NoSQL em empresas brasileiras, em especial as de pequeno/médio porte. Esse artigo tem como objetivo apresentar o resultado de um levantamento realizado com empresas brasileiras sobre a adoção de SGBDs NoSQL em suas atividades.

**Abstract.** In recent years, an increase in the popularity of NoSQL Database Management Systems (DBMS) has been observed. These DBMSs commonly offer flexibility in data representation and facilitated horizontal scalability. Although it is common knowledge that large international companies adopt these solutions, the scenario of NoSQL DBMS adoption in Brazilian companies, especially small/medium-sized ones, is not clear. This paper aims to present the results of a survey with Brazilian companies on the adoption of NoSQL DBMSs in their activities.

## 1. Introdução

Os Sistemas de Gerência de Banco de Dados Relacionais (SGBDRs) dominaram tanto a academia quanto a indústria nas últimas décadas. Uma vez que seguem o modelo relacional, tais sistemas são capazes de oferecer diversas vantagens, como o apoio para transações ACID e estruturas de indexação eficientes [Liu and Özsu 2018]. Apesar da adoção massiva de SGBDRs, sempre houve questionamentos acerca da sua adequabilidade para muitas aplicações (e.g., Redes Sociais), ou por conta de limitações de escalabilidade ou de apoio a tipos de dados heterogêneos (e.g., semiestruturados e não estruturados) [Stonebraker 2018].

Recentemente, temos visto um aumento considerável na adoção de SGBDs NoSQL como alternativa para esse cenários [Sadalage and Fowler 2012], especialmente em aplicações de Ciência de Dados. Existem diversos tipos de SGBDs NoSQL como os de Chave-Valor (e.g., Dynamo), orientados a documentos (e.g., MongoDB), colunares (e.g., Cassandra) e orientados a grafos (e.g., Neo4J). Cada um desses tipos de SGBD apresenta vantagens e desvantagens. Além disso, ainda existem soluções NoSQL de alta escalabilidade em nuvem, o que fomenta a adoção de tais soluções. Embora existam trabalhos anteriores que investigaram a adoção de SGBDs NoSQL em diversos cenários [Vera-Olivera et al. 2021, Ferrari et al. 2020, Davoudian et al. 2018] bem como relatos públicos de grandes empresas

---

\*O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. A pesquisa foi apoiada por CNPq e FAPERJ.

que já adotaram SGBDs NoSQL em suas atividades (*e.g.*, o Google com o BigTable, que é oferecido como um serviço na *Google Cloud Platform*, o Facebook com o Cassandra e o grupo Globo que adota o MongoDB aliado ao Redis), pouco se sabe sobre a adoção de SGBDs NoSQL em empresas brasileiras, especialmente as que atuam local/regionalmente.

O presente artigo tem como objetivo apresentar os resultados de um levantamento realizado com empresas brasileiras sobre a adoção de SGBDs NoSQL em suas atividades, com foco principal em empresas de pequeno/médio porte. A partir dos resultados coletados, pretendemos caracterizar o cenário atual para *a posteriori* identificar fatores importantes que podem fomentar (ou não) a adoção de tais sistemas no contexto brasileiro. Esse artigo se encontra organizado em quatro seções além da Introdução. A Seção 2 discute com um pouco mais de detalhes os SGBDs NoSQL existentes. A Seção 3 apresenta a metodologia adotada no levantamento. A Seção 4 discute os resultados, e, finalmente, a Seção 5 conclui o artigo e discute perspectivas futuras.

## 2. Referencial Teórico

Diversos SGBDs NoSQL têm sido propostos e implementados nos últimos anos, e diferentemente do esperado para SGBDRs, cada um segue um modelo de dados diferente [Corbellini et al. 2017]. Apesar desses SGBDs NoSQL possuírem diversas diferenças, algumas características são encontradas na maioria deles [Diogo et al. 2019]: *(i)* adoção de modelos de dados mais flexíveis, *(ii)* relaxamento das propriedades ACID, *(iii)* propensão para escalabilidade horizontal e *(iv)* facilidade para replicação e particionamento dos dados. Os SGBDs NoSQL podem ser classificados em quatro classes principais: *(i)* Chave-Valor, *(ii)* Orientados a Documentos, *(iii)* Colunares e *(iv)* Orientados a Grafos. A seguir discutimos cada uma dessas classes.

Os SGBDs chave-valor (*i.e.*, *key-value*) seguem um modelo de dados muito simples. Os dados são armazenados como pares chave-valor, onde cada par é identificado pelo valor de sua chave única. Apesar de possuir um modelo bastante simples, esse tipo de SGBD é utilizado em diversos cenários que variam desde armazenamento de dados de sessões de usuário na *Web* até dados relativos a jogos *multiplayer online*. Uma desvantagem desse tipo de SGBD é que a semântica dos dados se encontra na aplicação, e não no SGBD. Os SGBDs orientados a documentos tem como objetivo armazenar e consultar dados representados no formato de documento. Comumente esses documentos se encontram no formato JSON, BSON (um superconjunto do formato JSON que inclui outros tipos de dados, principalmente a matriz de *bytes*) e XML. Como muitas aplicações já manipulam dados em formatos semiestruturados, esse tipo de SGBD se mostra uma escolha natural para esses domínios de aplicações. Os SGBDs orientados a documentos não possuem um *schema* pré-definido, o que faz com que eles ofereçam um alto nível de flexibilidade.

Já os SGBDs colunares possuem uma estrutura lógica similar aos SGBDRs, uma vez que os dados também são organizados em relações. Entretanto, diferentemente dos SGBDRs, os SGBD colunares organizam seus dados no disco de acordo com as colunas e não com as tuplas de uma tabela. Esse tipo de SGBD se mostra bastante indicado para buscas que realizam múltiplas agregações em determinadas colunas. Por fim, os SGBDs orientados a grafos permitem que o usuário gerencie seus dados a partir de uma representação dos mesmos em forma de grafo. Os dados no banco de dados são modelados como vértices e arestas dos grafos. Cada vértice e aresta podem possuir atributos e as arestas possuem significância direcional, o que faz com que seja possível identificar padrões entre os nós. Apesar de útil para aplicações que naturalmente já tem seus dados em estruturas de grafo (*e.g.*, dados de

Proveniência), a adoção desse tipo de SGBD pode ser complexa em outros casos pois não há uma metodologia clara de como modelar o grafo. Assim, dependendo da consulta, os SGBDs de grafo podem apresentar uma perda de desempenho não-negligenciável.

### 3. Metodologia

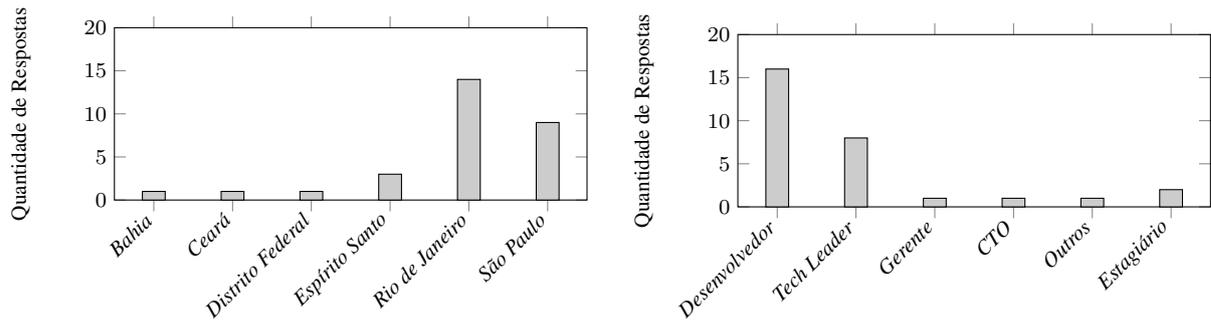
Esta seção discute a metodologia utilizada na construção de um questionário para compreender a adoção de SGBDs NoSQL em empresas brasileiras, especialmente as de pequeno/médio porte. Para sermos capazes de entender melhor o cenário, elaboramos um questionário e os disponibilizamos de forma *online*. O questionário foi distribuído em listas de *e-mail* e também diretamente para representantes de empresas e da academia. No projeto do questionário, buscou-se reduzir ao máximo as desistências, reduzindo o esforço do participante. Assim, quase todas as questões apresentavam opções pré-definidas para escolha. Algumas das perguntas permitiram que os participantes selecionassem múltiplas respostas, enquanto a maioria se restringiu a apenas uma resposta. O participante podia apenas responder parte do questionário, caso desejasse. No total, 76 participantes responderam pelo menos uma pergunta do questionário no período de 16 de novembro de 2021 até 20 de março de 2022, sendo que 29 participantes responderam a todas as perguntas. Em nossas análises abaixo, consideramos apenas essa população de 29 participantes.

Na primeira parte do questionário, definimos como objetivo caracterizar os participantes e as empresas onde trabalham. Portanto, foram elaboradas e respondidas as seguintes questões: (i) Em que estado sua empresa se localiza? (ii) Qual seu cargo/função na empresa? (iii) Qual o setor de atuação da empresa? e (iv) Qual o número de funcionários da empresa? Na segunda parte do questionário, definimos como objetivo entender o cenário atual de adoção de SGBDs NoSQL nessas empresas. Para tanto, definimos sete questões de pesquisa: (i) QP1. O uso de SGBDs NoSQL reduz as despesas de TI na empresa? (ii) QP2. O uso de SGBDs NoSQL oferece novas oportunidades de negócio? (iii) QP3. O uso de SGBDs NoSQL oferece confiabilidade no uso? (iv) QP4. O uso de SGBDs NoSQL oferece maior disponibilidade no serviço? (v) QP5. O uso de SGBDs NoSQL oferece maior segurança no serviço? (vi) QP6. A adoção de SGBDs NoSQL introduziu dificuldades de integração com os sistemas existentes na empresa? e (vii) QP7. Qual SGBD NoSQL a empresa usa no momento? Cada questão de pesquisa foi respondida seguindo a escala de *Likert* com as seguintes opções: 1) DT – Discordo Totalmente, 2) DP – Discordo Parcialmente, 3) SO – Sem Opinião Formada, 4) CP – Concordo Parcialmente e 5) CT – Concordo Totalmente. A Seção 4 analisa as respostas obtidas de acordo com o questionário elaborado.

### 4. Análise dos Resultados

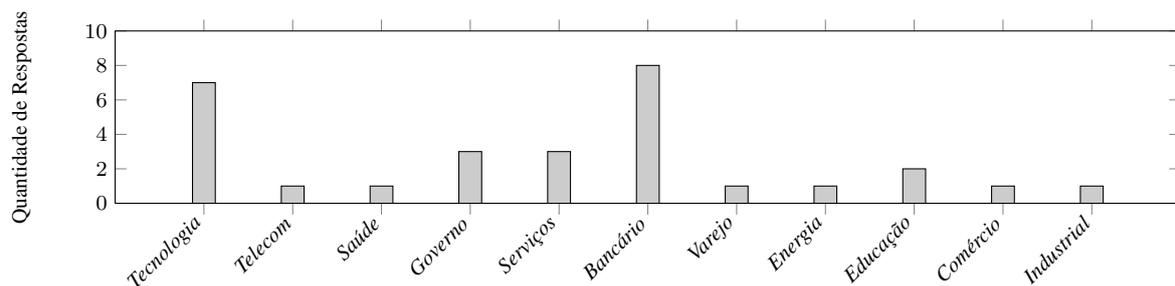
Nessa seção analisamos os resultados obtidos com o questionário apresentado na seção anterior. A primeira parte da Figura 1 apresenta a distribuição geográfica das empresas. A grande maioria dos participantes trabalham em empresas na região sudeste (em especial RJ e SP). Em termos de cargo ocupado dentro da empresa, a maioria dos participantes é desenvolvedor de *software* ou *Tech Leader*, o que demonstra uma certa experiência no uso de tecnologias, conforme pode ser visualizado na segunda parte da Figura 1.

Além disso, analisamos também o setor de atuação da empresa, conforme pode ser visualizado na Figura 2. Conforme esperado, uma grande parcela das empresas atua no setor de tecnologia. Porém, um resultado surpreendente foi a quantidade de empresas do setor bancário. Por necessidade, regulação e custos de negócio, o setor bancário possui diversos sistemas legados, o que, em teoria, faria que a adoção de novas tecnologias como SGBDs

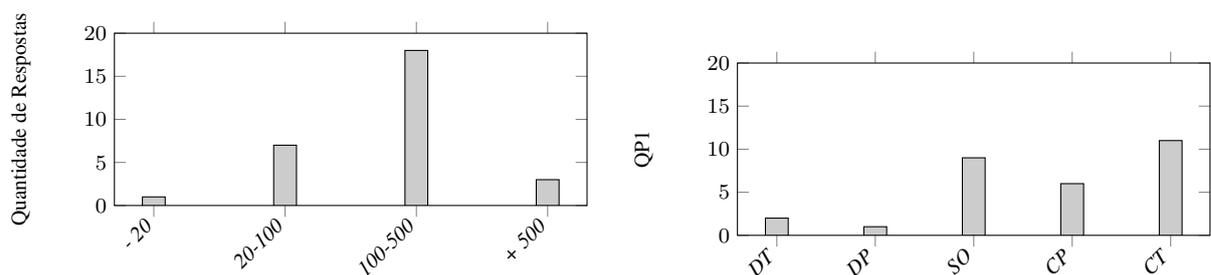


**Figura 1. Localização da empresa por estado da federação. Cargo do participante na empresa no momento da resposta do questionário.**

NoSQL fosse mais lenta. Acreditamos que tal adoção se deva ao aparecimento de diversas *Fintechs* nos últimos anos, que tendem a adotar tecnologias mais recentes. Finalmente, na Figura 3(a), analisamos as empresas de acordo com a quantidade de colaboradores. A maioria dos participantes trabalha em empresas de pequeno (até 100 colaboradores) e médio porte (entre 100 e 500 colaboradores). Essa distribuição era justamente a planejada, pois pretendemos analisar se a adoção de SGBDs NoSQL nessas empresas difere do padrão das grandes empresas como discutido em [Davoudian et al. 2018].



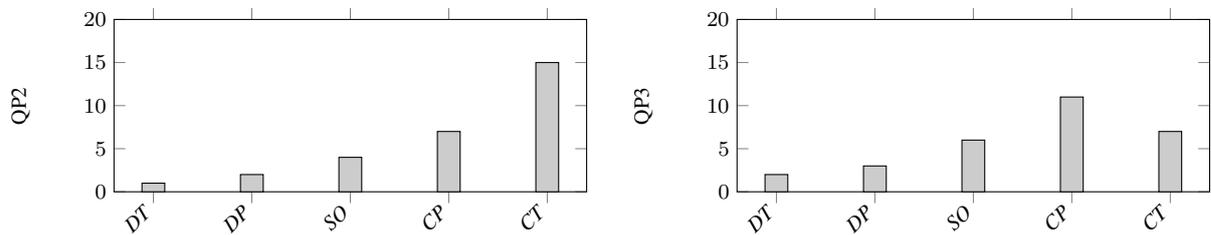
**Figura 2. Setor de Atuação da Empresa**



**Figura 3. Número de colaboradores. Resposta para a QP1 – O uso de SGBD NoSQL reduz as despesas de TI na empresa?**

Para responder às questões de pesquisa (QP1 a QP7) analisemos os resultados apresentados nas Figuras 3 a Figura 6. Em relação à QP1 (Figura 3), podemos perceber que a maioria dos participantes acredita que a adoção de SGBDs NoSQL reduz despesas de TI na empresa. Em especial, os *Tech Leaders* informaram que o fato de muitos SGBDs NoSQL serem *Open Source* e gratuitos impacta nessa redução. Em relação à QP2 e QP3 (Figura 4)), podemos perceber que a adoção de SGBDs NoSQL ofereceu novas oportunidades no

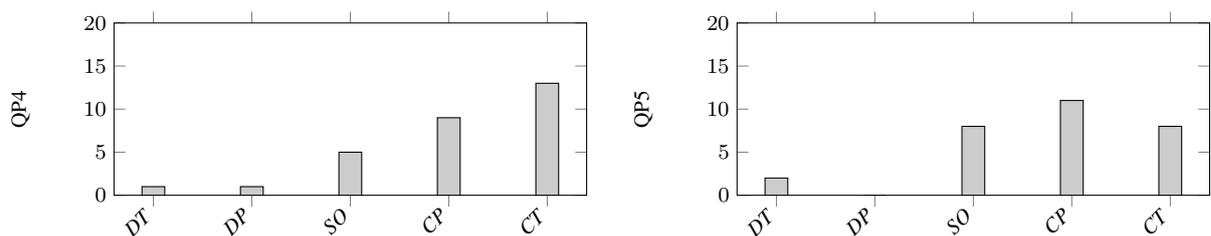
negócio, mas que o aumento na confiabilidade não se mostra tão claro para a maioria dos participantes, se comparado com o uso de SGBDRs.



**Figura 4. QP2 – O uso de SGBD NoSQL ofereceu novas oportunidades de negócio? QP3 – O uso de SGBD NoSQL ofereceu confiabilidade no serviço?**

Entretanto, ao analisarmos a QP4 e QP5 (Figura 5) podemos perceber que a clara maioria dos participantes concorda que a adoção de SGBDs NoSQL aumentou a disponibilidade e segurança do serviço. Esse resultado é bastante compreensível dado que muitas das empresas utilizam serviços de nuvem que oferecem SGBDs NoSQL, *e.g.*, *Google Cloud Datastore*, que oferecem disponibilidade do serviço de 99.99%. Apesar de enxergarem vantagens na adoção de SGBDs NoSQL, os participantes também relataram dificuldades em integrar os sistemas existentes com soluções NoSQL (Figura 6). Alguns participantes mencionaram que em alguns casos a integração com sistemas legados (*e.g.*, implementados na linguagem de programação Natural e usando ADABAS como solução de banco de dados) exigiu a implementação de complexos componentes mediadores.

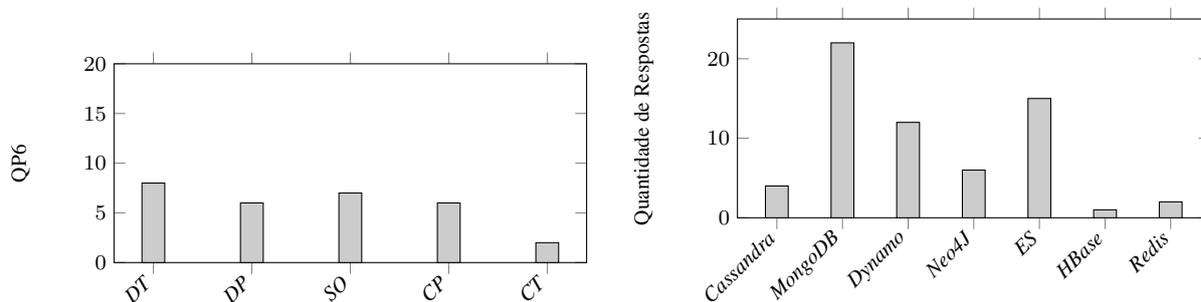
Finalmente, a Figura 6 também apresenta os SGBD NoSQL adotados nas empresas. Podemos perceber que o MongoDB, o Dynamo e o ElastiSearch dominam as soluções adotadas. Em especial, os SGBDs orientados a documentos e Chave-Valor tiveram sua adoção justificada a partir da necessidade das empresas gerenciarem dados que já se encontram semi-estruturados (*e.g.*, arquivos XML e JSON extraídos a partir de serviços públicos como NFe no caso das empresas da área do varejo) ou em aplicações que precisam armazenar dados com estruturas mais simples (como controles de sessão do usuário em aplicações desenvolvidas pelas empresas da área de educação).



**Figura 5. QP4 – O uso de SGBD NoSQL ofereceu disponibilidade no serviço? QP5 – O uso de SGBD NoSQL ofereceu segurança no serviço?**

Apesar das respostas indicarem uma leve tendência de adoção de SGBDs NoSQL em empresas de pequeno-médio porte no Brasil, tal conclusão ainda não pode ser generalizada, uma vez que a quantidade de respostas obtidas ainda se mostra reduzida. Além disso, os resultados obtidos possuem algumas ameaças à validação. A principal ameaça à *validade interna* é a possibilidade de viés de seleção, pois compartilhamos o questionário com grupos próximos dos autores. Para mitigar essa ameaça, solicitamos às pessoas que encaminhassem o questionário para outros grupos. A representatividade dos participantes é a principal ameaça

à *validade externa*. Embora tenhamos recebido respostas de diferentes empresas, não foi possível garantir a visualização do questionário por empresas de todas as regiões do Brasil. Essa questão fica evidente dado o maior número de entrevistados da região sudeste.



**Figura 6. QP6 – O uso de SGBD NoSQL apresentou dificuldade ao integrar com os atuais sistemas da empresa? Qual SGBD NoSQL a empresa usa no momento?**

## 5. Conclusão

Esse artigo discutiu o fenômeno do aumento de popularidade e uso de SGBDs NoSQL no contexto de empresas locais/regionais brasileiras. Diferentemente de SGBDRs, os SGBDs NoSQL seguem diferentes modelos, o que faz com que sejam apropriados para determinados tipos de aplicação, mas não para outras. Apesar de ser de conhecimento público que grandes empresas internacionais utilizam esse tipo de SGBD, ainda não havia indícios sobre o cenário de adoção em empresas de pequeno/médio porte no Brasil. Assim, realizamos um levantamento inicial com empresas brasileiras para compreender a visão das mesmas sobre a adoção de SGBDs NoSQL. Os resultados mostraram que a grande maioria dos entrevistados considera que existem vantagens na adoção de SGBDs NoSQL nas empresas, mas que ainda existem desafios em sua adoção, principalmente no que tange a integração com sistemas existentes. Trabalhos futuros incluem a execução do levantamento com um número maior de empresas, uma vez que a amostra apresentada nesse artigo ainda é reduzida.

## Referências

- Corbellini, A., Mateos, C., Zunino, A., Godoy, D., and Schiaffino, S. (2017). Persisting big-data: The nosql landscape. *Information Systems*, 63:1–23.
- Davoudian, A., Chen, L., and Liu, M. (2018). A survey on nosql stores. *ACM Comp. Surv.*, 51(2).
- Diogo, M., Cabral, B., and Bernardino, J. (2019). Consistency models of nosql databases. *Future Internet*, 11(2).
- Ferrari, D., Carminati, M., Polino, M., and Zanero, S. (2020). Nosql breakdown: A large-scale analysis of misconfigured nosql services. *ACSAC '20*, page 567–581.
- Liu, L. and Özsu, M. T., editors (2018). *Encyclopedia of Database Systems*. Springer.
- Sadalage, P. J. and Fowler, M. (2012). *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*. Addison-Wesley Professional, 1st edition.
- Stonebraker, M., editor (2018). *Making Databases Work: The Pragmatic Wisdom of Michael Stonebraker*, volume 22. ACM and Morgan & Claypool.
- Vera-Olivera, H., Guo, R., Bernardi, A. P., and Holanda, M. (2021). Data modeling and nosql databases - a systematic mapping review. *ACM Comput. Surv.*, 54(6).