

# Análise Experimental de Tecnologias de Banco de Dados para Aplicações em Tempo Real no Futebol

Rael Kiluanji de Jesus Cassimiro<sup>1</sup>

<sup>1</sup>Instituto de Ciências Exatas e Informática – Pontifícia Universidade Católica de Minas Gerais  
(PUC MINAS)

Caixa Postal, 1.686 CEP 30535-901 – Belo Horizonte – Minas Gerais – Brasil

raelkiluanji@gmail.com, rael.cassimiro@sga.pucminas.br

**Abstract.** *This paper presents an experimental analysis of database technologies applied to real-time sports data analysis. The objective is to identify which solutions deliver better technical performance in high-demand scenarios, considering metrics such as latency, scalability, and resource consumption. The research adopts an applied, quantitative, and experimental approach, performing controlled tests with different database management systems in simulated professional football scenarios. The results support the categorization of the evaluated solutions and provide guidelines for selecting appropriate technologies for real-time sports applications.*

**Resumo.** *Este artigo apresenta uma análise experimental de tecnologias de banco de dados aplicadas à análise de dados esportivos em tempo real. O objetivo é identificar quais soluções oferecem melhor desempenho técnico em contextos de uso intensivo, considerando métricas como latência, escalabilidade e consumo de recursos. A pesquisa adota abordagem aplicada, quantitativa e experimental, conduzindo testes controlados com diferentes sistemas de gerenciamento de banco de dados em cenários simulados do futebol profissional. Os resultados permitem categorizar as soluções avaliadas e propor diretrizes para escolha tecnológica em aplicações esportivas com exigência de processamento em tempo real.*

**Bacharelado em Engenharia de Software - PUC Minas**  
**Trabalho de Conclusão de Curso (TCC)**

Orientador de conteúdo (TCC I): Joana Souza - joanasouza@pucminas.br

Orientador de conteúdo (TCC I): João Pedro Batisteli - jp.batisteli@hotmail.com

Orientador de conteúdo (TCC I): Leonardo Vilela - leonardocardoso@pucminas.br

Orientador acadêmico (TCC I): Cleiton Tavares - cleitontavares@pucminas.br

Orientador do TCC II: (A ser definido no próximo semestre)

Belo Horizonte, 18 de 05 de 2025.

## 1. Introdução

A análise de dados em tempo real tem ganhado destaque como estratégia essencial para apoiar decisões em diversos setores, incluindo o futebol profissional. O uso crescente

de sensores embarcados, dispositivos vestíveis e plataformas de transmissão digital viabiliza a coleta contínua de dados durante partidas e treinamentos. Esses dados permitem ações como monitoramento de desempenho, prevenção de lesões e suporte a decisões táticas. Ponciano et al. [Ponciano et al. 2023] ilustram a aplicação de algoritmos de aprendizado de máquina na previsão de resultados em competições esportivas. Valadão et al. [Valadão et al. 2023] exploram o uso de dados operacionais para caracterizar estratégias de jogo, com base na frequência e efetividade de jogadas. Já Calmon et al. [Calmon et al. 2024] demonstram o potencial da integração de diferentes bases de dados na identificação de atletas promissores, indicando a importância de soluções eficientes de armazenamento e processamento.

Apesar dos avanços na coleta e armazenamento de dados esportivos, o uso analítico em tempo real ainda enfrenta desafios técnicos consideráveis. Limitações como alta latência nas consultas, consumo excessivo de recursos e baixa escalabilidade dificultam a aplicação prática desses dados em cenários de decisão rápida. Pesquisas recentes apontam que o desempenho das tecnologias de banco de dados varia significativamente conforme o contexto e a carga de trabalho [Rahman et al. 2024, Tan e Zhang 2024], demandando análises experimentais criteriosas. **O problema que esta pesquisa busca investigar é: quais tecnologias de banco de dados oferecem melhor desempenho técnico em contextos de análise de dados esportivos em tempo real, considerando métricas como tempo de resposta, escalabilidade e eficiência no uso de recursos?**

A relevância do problema proposto reside na crescente demanda por soluções que processem dados esportivos de forma eficiente, confiável e escalável. A análise precisa desses dados pode impactar diretamente o desempenho estratégico de equipes profissionais. Valadão et al. [Valadão et al. 2023] evidenciam que estratégias de jogo podem ser caracterizadas com base na frequência e efetividade das jogadas, desde que os dados sejam devidamente processados. Já Calmon et al. [Calmon et al. 2024] demonstram que a integração de múltiplas bases e a aplicação de algoritmos de aprendizado de máquina podem prever o sucesso de jovens atletas, o que depende de infraestruturas de dados robustas. Essas evidências reforçam a necessidade de pesquisas que avaliem, sob condições controladas, o desempenho de diferentes tecnologias de banco de dados frente às exigências do cenário esportivo profissional.

**O objetivo geral deste trabalho é analisar criticamente e categorizar soluções de banco de dados aplicadas à análise de dados esportivos em tempo real no domínio do futebol.** Especificamente, busca-se identificar os principais requisitos técnicos relacionados ao desempenho e à escalabilidade que impactam a eficiência dessas tecnologias; examinar alternativas disponíveis na literatura sob a perspectiva da latência, confiabilidade e integração com sistemas analíticos; além de comparar essas soluções por meio de experimentos controlados, propondo diretrizes que orientem a seleção de tecnologias adequadas ao contexto futebolístico.

Como resultado, pretende-se apresentar uma categorização crítica das soluções de banco de dados para análise de dados futebolísticos em tempo real, considerando aspectos como desempenho em consultas, escalabilidade, confiabilidade e integração com sistemas analíticos.

O restante deste trabalho está organizado em três seções, além desta introdução.

A Seção 2 aborda a fundamentação teórica, enquanto a Seção 3 trata dos trabalhos relacionados. A Seção 4 detalha os materiais e métodos adotados.

## 2. Fundamentação Teórica

### 2.1. Images

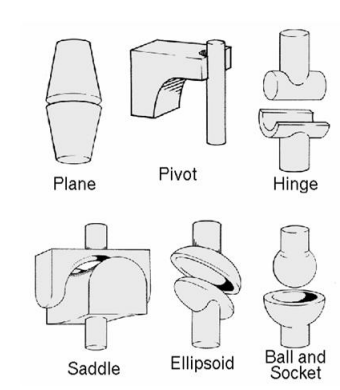
All images and illustrations should be in black-and-white, or gray tones, excepting for the papers that will be electronically available (on CD-ROMs, internet, etc.). The image resolution on paper should be about 600 dpi for black-and-white images, and 150-300 dpi for grayscale images. Do not include images with excessive resolution, as they may take hours to print, without any visible difference in the result.

### 2.2. Figures and Captions

Figure and table captions should be centered if less than one line (Figure 1), otherwise justified and indented by 0.8cm on both margins, as shown in Figure 2. The caption font must be Helvetica, 10 point, boldface, with 6 points of space before and after each caption.



**Figura 1. A typical figure**



**Figura 2. This figure is an example of a figure caption taking more than one line and justified considering margins mentioned in Section 2.2.**

### 3. Trabalhos Relacionados

Os trabalhos relacionados discutidos nesta seção envolvem o uso de bancos de dados e técnicas de análise de dados em tempo real, com ênfase em aplicações no domínio esportivo e em avaliações empíricas de desempenho e eficiência. Os estudos selecionados também abordam o uso de aprendizado de máquina em contextos similares. Dessa forma, os trabalhos estão organizados de acordo com o seu nível de aproximação com a proposta desta pesquisa, considerando tanto o objeto de estudo quanto a abordagem metodológica adotada.

Suh et al. [Suh et al. 2022] realizaram um estudo empírico abrangente sobre o desempenho de consultas em sistemas de banco de dados baseados em GPU. Foram avaliados diferentes SGBDs utilizando benchmarks padronizados, considerando operações complexas como JOIN e agregações. Os autores empregaram um modelo causal para identificar os fatores mais impactantes no tempo de execução das consultas, revelando gargalos como a transferência de dados entre CPU e GPU e o uso limitado de memória da GPU. Este trabalho contribui para a presente pesquisa ao oferecer uma base metodológica sólida para comparação de desempenho entre SGBDs, especialmente úteis em aplicações de análise em tempo real, como o domínio esportivo explorado neste projeto.

Lella et al. [Lella et al. 2024] investigaram o consumo energético de sistemas de gerenciamento de banco de dados SQL e NoSQL, propondo uma ferramenta chamada DBJoules. O estudo avaliou operações típicas como SELECT, INSERT, UPDATE, DELETE e JOIN, em bancos como MySQL, PostgreSQL e MongoDB. Os resultados indicaram variações significativas no consumo de energia conforme a operação e a tecnologia utilizada. A presente pesquisa poderá se beneficiar dessas análises ao considerar aspectos de eficiência e sustentabilidade na escolha do SGBD mais adequado para aplicações esportivas com processamento em tempo real.

Ponciano et al. [Ponciano et al. 2023] exploraram o uso de algoritmos de aprendizado de máquina para prever os vencedores dos playoffs da NFL. A abordagem envolveu coleta e preparação de dados históricos, aplicação de modelos como Random Forest e redes neurais artificiais, além da avaliação com métricas como F1-score e precisão. Os resultados demonstraram bom desempenho dos modelos e destacaram a importância de fatores como penalidades e jogadas ofensivas. Este trabalho serve como referência direta para a presente pesquisa, tanto na estrutura metodológica quanto na validação do uso de modelos de AM na previsão de eventos esportivos.

Calmon et al. [Calmon et al. 2024] propuseram um modelo para prever o sucesso de jovens atletas de futebol com base em dados estatísticos. Foram utilizados algoritmos de classificação e seleção de atributos relevantes, visando identificar padrões associados a transições bem-sucedidas para níveis profissionais. A metodologia e o domínio de aplicação se aproximam diretamente da proposta desta pesquisa, fornecendo subsídios para estruturação de modelos preditivos voltados ao desempenho em jogos e análise de atletas.

Por fim, Souza e Oliveira [Souza e Oliveira 2023] compararam os serviços Firebase e PostgreSQL com foco em critérios como desempenho, facilidade de uso e escalabilidade. Embora não seja um estudo empírico clássico, a análise técnica oferecida pelos autores é útil para fundamentar escolhas de infraestrutura de armazenamento em proje-

tos com exigências específicas de resposta em tempo real. Esta comparação poderá ser considerada na definição da arquitetura do sistema proposto neste trabalho.

## 4. Materiais e Métodos

A definição dos materiais e métodos empregados neste estudo está alinhada aos objetivos delineados na introdução e fundamentada nas contribuições discutidas nos trabalhos relacionados. A abordagem metodológica busca simular cenários típicos do domínio do futebol profissional, de modo a possibilitar a medição de desempenho de tecnologias de banco de dados sob condições equivalentes de uso.

### 4.1. Tipo de Pesquisa

Este trabalho configura-se como uma **pesquisa aplicada**, voltada à resolução de problemas práticos relacionados à eficiência de tecnologias de banco de dados em aplicações esportivas. Trata-se também de um estudo **experimental**, pois envolve a execução de testes controlados em ambiente simulado, com o objetivo de mensurar o desempenho de diferentes tecnologias de gerenciamento de banco de dados.

A definição metodológica deste estudo foi inspirada em abordagens previamente adotadas em trabalhos da área. Suh et al. [Suh et al. 2022] conduziram experimentos com SGBDs em ambientes de GPU para medir desempenho; Lella et al. [Lella et al. 2024] analisaram o consumo energético em consultas a bancos de dados; Rahman et al. [Rahman et al. 2024] estudaram estratégias de otimização de SQL para grandes volumes de dados; e Tan e Zhang [Tan e Zhang 2024] exploraram o desempenho de consultas em pipelines automatizados de bancos de dados.

### 4.2. Ambiente de Testes

Os experimentos desta pesquisa serão conduzidos em um ambiente híbrido, composto por infraestrutura local e serviços de computação em nuvem (AWS e Google Cloud), a fim de simular diferentes cenários operacionais. As configurações foram selecionadas com base na literatura especializada e na viabilidade de replicação dos testes.

- **Sistemas de gerenciamento de banco de dados (SGBDs):** serão utilizados PostgreSQL (modelo relacional), MongoDB (NoSQL orientado a documentos) e Firebase (banco em tempo real). O uso do PostgreSQL se justifica por sua robustez em consultas transacionais e análises estruturadas [Souza e Oliveira 2023], enquanto o MongoDB destaca-se por sua flexibilidade no armazenamento de dados semi-estruturados, comuns em aplicações esportivas. Já o Firebase é amplamente adotado em aplicações com requisitos de atualização em tempo real, como demonstrado por Calmon et al. [Calmon et al. 2024].
- **Sistema operacional:** Ubuntu Server 22.04 LTS, escolhido por sua estabilidade, suporte a ambientes de teste e compatibilidade com as ferramentas adotadas [Rahman et al. 2024].
- **Linguagem de programação e bibliotecas:** Python 3.11, com uso das bibliotecas `psycopg2` (PostgreSQL), `pymongo` (MongoDB), `sqlalchemy` (ORM), entre outras para automação dos testes e análise estatística.
- **Infraestrutura de hardware:** máquinas virtuais com configurações mínimas de 4 vCPUs, 8 GB de RAM e armazenamento SSD de 256 GB, tanto localmente quanto em instâncias de nuvem, garantindo homogeneidade e reprodutibilidade.

A seleção desses elementos considera a necessidade de avaliar o comportamento das tecnologias em condições realistas, conforme práticas observadas em estudos anteriores que analisam desempenho, escalabilidade e custo operacional em ambientes diversos.

### 4.3. Métodos Utilizados

A metodologia adotada baseia-se na execução de **experimentos controlados**, organizados em uma sequência estruturada de etapas, com o objetivo de simular cenários típicos do uso de SGBDs em aplicações esportivas. Os testes são orientados por quatro eixos metodológicos, com base em estudos prévios e práticas consolidadas na área.

1. **Definição dos cenários de consulta:** inspirados em padrões de uso identificados por Tan e Zhang [Tan e Zhang 2024] e Valadão et al. [Valadão et al. 2023], serão modelados diferentes tipos de consultas, como agregações, filtros por tempo e junções entre entidades esportivas.
2. **Preparação das bases de dados:** serão utilizados conjuntos de dados públicos e simulados, com volumes de 10 mil, 100 mil e 1 milhão de registros, representando situações de baixa, média e alta carga. Essa variação permite testar a **escalabilidade dos sistemas** e segue a estratégia de amostragem por conveniência, adotada por Calmon et al. [Calmon et al. 2024].
3. **Execução dos testes comparativos:** as tecnologias PostgreSQL, MongoDB e Firebase serão submetidas a cargas idênticas em ambientes equivalentes. A estrutura de testes será baseada no protocolo experimental descrito por Souza e Oliveira [Souza e Oliveira 2023], garantindo reprodutibilidade e controle de variáveis externas.
4. **Coleta e análise dos resultados:** os dados obtidos (tempo de resposta, taxa de falha, consumo energético) serão analisados por meio de estatísticas descritivas (média, desvio padrão e intervalo de confiança), com apoio das bibliotecas Pandas e SciPy. Essa abordagem segue práticas adotadas por Lella et al. [Lella et al. 2024] e Rahman et al. [Rahman et al. 2024].

Todos os procedimentos serão registrados em Jupyter Notebooks e versionados em repositório Git, garantindo transparência e reprodutibilidade. Não há aplicação de algoritmos de mineração de dados nesta etapa, pois o foco está na avaliação de desempenho técnico dos SGBDs sob diferentes condições de carga.

### 4.4. Métricas de Avaliação

Para a avaliação das tecnologias testadas, serão consideradas métricas técnicas amplamente utilizadas em estudos sobre desempenho de bancos de dados. Cada métrica foi selecionada por sua relevância em contextos de análise esportiva em tempo real e será coletada por meio de ferramentas automatizadas e scripts de monitoramento.

- **Tempo de resposta (ms)** — mede a latência média por tipo de consulta, permitindo avaliar a eficiência do sistema sob diferentes cargas. Essa métrica é central em estudos como o de Rahman et al. [Rahman et al. 2024].
- **Taxa de sucesso (%)** — indica a proporção de consultas executadas com sucesso, sendo essencial para avaliar a confiabilidade das tecnologias. Utilizada por Ponciano et al. [Ponciano et al. 2023] em ambientes de análise preditiva.

- **Consumo energético (estimado)** — obtido com auxílio da ferramenta DBJoules, que permite estimar o impacto energético de cada operação, conforme metodologia de Lella et al. [Lella et al. 2024].
- **Custo operacional (por mil requisições)** — estimado a partir da simulação em serviços de nuvem (AWS e Google Cloud), com base na medição de tempo e recursos utilizados por operação.
- **Escalabilidade** — avaliada pela variação do tempo de resposta e consumo à medida que o volume de dados aumenta, refletindo a capacidade do sistema de lidar com grandes quantidades de informação. Essa análise está alinhada à abordagem de Neto e Silva [Neto e Silva 2023].

Cada métrica será analisada estatisticamente com base em valores médios, desvio padrão e gráficos de dispersão. As medições ocorrerão após a execução de cada cenário experimental descrito na subseção anterior, garantindo consistência e comparabilidade dos resultados.

### 4.5. Instrumentos Utilizados

Para a condução dos testes, serão empregados:

- Jupyter Notebooks para orquestração dos testes e visualização de resultados;
- DBJoules, conforme descrito por Lella et al. [Lella et al. 2024], para mensuração de consumo;
- pgAdmin e Firebase Console para administração dos SGBDs;
- Scripts automatizados de carga e medição baseados em Python.

Não estão previstos questionários nem entrevistas, visto que o foco é técnico-computacional, e todos os dados utilizados são públicos ou simulados.

### 4.6. Cronograma de Execução

A Tabela 1 apresenta o cronograma quinzenal das atividades previstas para a execução do TCC II, abrangendo o período de agosto a dezembro de 2025. As etapas foram organizadas de forma sequencial, considerando a dependência entre tarefas e a alocação de tempo adequada para testes, análise e escrita.

Tabela 1. Cronograma das atividades previstas (agosto a dezembro de 2025)

Etapa	Ago-1	Ago-2	Set-1	Set-2	Out-1	Out-2	Nov
A1 – Levantamento de requisitos técnicos	X	X					
A2 – Configuração do ambiente de testes		X	X				
A3 – Implementação dos scripts de benchmark			X	X			
A4 – Execução dos experimentos				X	X	X	
A5 – Coleta e análise estatística dos dados					X	X	X
A6 – Elaboração da redação técnica (parcial)						X	X
A7 – Revisão metodológica e ajustes							
A8 – Escrita e finalização do TCC II							

## Referências

- CALMON, L. et al. Previsão de sucesso de atletas jovens de futebol usando integração de diferentes base de dados. In: *Anais do XXXIX Simpósio Brasileiro de Bancos de Dados (SBBD 2024)*. [S.l.: s.n.], 2024.
- LELLA, H. S. et al. Towards comprehending energy consumption of database management systems: A tool and empirical study. In: *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering (EASE)*. [S.l.: s.n.], 2024.
- NETO, J. dos S.; SILVA, A. A. Estudo do impacto da adoção de tecnologias digitais no futebol. *RECIMA21 - Revista Científica Multidisciplinar*, v. 4, n. 3, 2023. Disponível em: <<https://doi.org/10.47820/recima21.v4i3.2923>>.
- PONCIANO, L. et al. Predizendo os vencedores dos playoffs: Um estudo de caso com aprendizado de máquina em partidas de futebol americano. In: *Proceedings of the 38th Brazilian Symposium on Databases (SBBD)*. [S.l.: s.n.], 2023. p. 1–7.
- RAHMAN, M. M. et al. Advanced query optimization in sql databases for real-time big data analytics. *Asian Journal of Business and Information Systems*, v. 4, n. 3, 2024. Disponível em: <<https://doi.org/10.69593/ajbais.v4i3.77>>.
- SOUZA, I.; OLIVEIRA, T. Comparação de serviços de banco de dados na nuvem: Amazon rds vs google cloud sql. In: *Anais do Simpósio Brasileiro de Sistemas de Informação (SBSI)*. [S.l.: s.n.], 2023. p. 1–8.
- SUH, A. R. et al. A comprehensive empirical study of query performance across gpudbm-ses. In: *Proceedings of the VLDB Endowment*. [S.l.: s.n.], 2022. v. 15, n. 9, p. 1842–1854.
- TAN, X.; ZHANG, Y. Research on automated database health check and slow sql optimization. In: *Proceedings of the 2024 International Conference on Cloud Computing and Big Data (ICCBD '24)*. [S.l.: s.n.], 2024. p. 294–299.
- VALADÃO, G. et al. Caracterização de estratégias de futebol com base na frequência, importância e efetividade de jogadas. In: *Anais do XX Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2023)*. [S.l.: s.n.], 2023.

## Referências

- CALMON, L. et al. Previsão de sucesso de atletas jovens de futebol usando integração de diferentes base de dados. In: *Anais do XXXIX Simpósio Brasileiro de Bancos de Dados (SBBD 2024)*. [S.l.: s.n.], 2024.
- LELLA, H. S. et al. Towards comprehending energy consumption of database management systems: A tool and empirical study. In: *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering (EASE)*. [S.l.: s.n.], 2024.
- NETO, J. dos S.; SILVA, A. A. Estudo do impacto da adoção de tecnologias digitais no futebol. *RECIMA21 - Revista Científica Multidisciplinar*, v. 4, n. 3, 2023. Disponível em: <<https://doi.org/10.47820/recima21.v4i3.2923>>.



- PONCIANO, L. et al. Predizendo os vencedores dos playoffs: Um estudo de caso com aprendizado de máquina em partidas de futebol americano. In: *Proceedings of the 38th Brazilian Symposium on Databases (SBBD)*. [S.l.: s.n.], 2023. p. 1–7.
- RAHMAN, M. M. et al. Advanced query optimization in sql databases for real-time big data analytics. *Asian Journal of Business and Information Systems*, v. 4, n. 3, 2024. Disponível em: <<https://doi.org/10.69593/ajbais.v4i3.77>>.
- SOUZA, I.; OLIVEIRA, T. Comparação de serviços de banco de dados na nuvem: Amazon rds vs google cloud sql. In: *Anais do Simpósio Brasileiro de Sistemas de Informação (SBSI)*. [S.l.: s.n.], 2023. p. 1–8.
- SUH, A. R. et al. A comprehensive empirical study of query performance across gpudbm-ses. In: *Proceedings of the VLDB Endowment*. [S.l.: s.n.], 2022. v. 15, n. 9, p. 1842–1854.
- TAN, X.; ZHANG, Y. Research on automated database health check and slow sql optimization. In: *Proceedings of the 2024 International Conference on Cloud Computing and Big Data (ICCBD '24)*. [S.l.: s.n.], 2024. p. 294–299.
- VALADÃO, G. et al. Caracterização de estratégias de futebol com base na frequência, importância e efetividade de jogadas. In: *Anais do XX Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2023)*. [S.l.: s.n.], 2023.

## **Apêndice**

### **A. Exemplo de Seção em um Apêndice**