

IMPLEMENTAÇÃO DE COMPUTAÇÃO DE BORDA PARA REDUÇÃO DE LATÊNCIA EM UM SISTEMA DE PRONTUÁRIO ELETRÔNICO PARA SPA'S

IMPLEMENTATION OF EDGE COMPUTING TO REDUCE LATENCY IN AN
ELECTRONIC HEALTH RECORD SYSTEM FOR SPAS

Ciências Exatas e da Terra • 24/04/2026

REGISTRO DOI: [10.70773/revistatopicos/776967150](https://doi.org/10.70773/revistatopicos/776967150)

Ailton Marques dos Santos Junior¹

Jean Mark Lobo de Oliveira²

Jean Carlos Araujo de Figueiredo³

Aguinaldo Alves Braga Neto⁴

RESUMO

Este estudo explora como a computação de borda (*edge computing*) pode ser a peça-chave para resolver o problema da lentidão em prontuários eletrônicos dentro de Unidades de Pronto Atendimento (SPA). Sabemos que depender apenas da nuvem em situações de emergência é arriscado, pois falhas de conexão ou gargalos digitais podem atrasar atendimentos vitais. Ao testarmos um cenário que simula o dia a dia real de uma unidade de saúde, os números mostraram uma mudança impressionante: a latência, que era de 210 ms na nuvem, despencou para apenas 32 ms com o processamento local. Mais do que velocidade, essa arquitetura provou ser extremamente resiliente, mantendo o sistema operando mesmo sem internet e garantindo que os dados dos pacientes fiquem protegidos e dentro das normas da LGPD. Em última análise, a computação de borda se consolida como uma estratégia essencial para modernizar a saúde digital, oferecendo o suporte rápido e seguro que médicos e pacientes precisam em momentos críticos.

Palavras-chave: Computação de Borda; Prontuário Eletrônico; Latência; Saúde Digital; Unidades de Pronto Atendimento.

ABSTRACT

This study explores how edge computing can be the key to solving latency issues in electronic health records within Urgent Care Units (SPA). Relying solely on the cloud in emergency situations is risky, as connection failures or digital bottlenecks can delay vital care. By testing a scenario that simulates the actual daily routine of a healthcare unit, the numbers showed a striking change: latency, which was 210 ms in the cloud, plummeted to just 32 ms with local processing. More than just speed, this architecture proved to be extremely resilient, keeping the¹ system running even without internet access and ensuring that patient data remains protected

and compliant with LGPD regulations. Ultimately, edge computing stands out as an essential strategy for modernizing digital health, providing the fast and secure support that doctors and patients need in critical moments.

Keywords: Edge Computing; Electronic Health Record; Latency; Digital Health; Urgent Care Units.

1. INTRODUÇÃO

No dinâmico cenário da saúde digital a agilidade no acesso aos dados clínicos em Unidades de Serviço de Pronto Atendimento (SPA) não é apenas uma meta tecnológica, mas um requisito vital, visto que a dependência exclusiva de servidores em nuvem pode gerar latências críticas que comprometem o tempo de resposta em situações de emergência. O prontuário eletrônico atua como o núcleo da inteligência assistencial, porém, sua eficiência é frequentemente limitada por gargalos de rede e instabilidades na conexão externa, o que torna a computação de borda (*edge computing*) uma solução estratégica ao processar dados na periferia da rede, garantindo que as informações cruciais estejam disponíveis instantaneamente para a equipe médica. Ao aproximar o poder de processamento do ponto de cuidado, essa arquitetura reduz drasticamente o tráfego de dados e assegura a continuidade operacional da unidade mesmo diante de falhas de conectividade global, transformando a infraestrutura de TI em um aliado resiliente para a sobrevivência do paciente e a fluidez do fluxo hospitalar, permitindo que a tecnologia acompanhe a urgência da vida real.

A implementação da computação de borda justifica-se pela necessidade premente de mitigar riscos operacionais e atrasos sistêmicos que impactam diretamente a segurança do paciente em

ambientes de alta criticidade. Segundo Rezende (2021), a gestão eficiente de dados em saúde é indissociável de uma infraestrutura que suporte o fluxo contínuo de informações sem interrupções, superando as limitações de banda que muitas vezes afetam modelos centralizados em nuvem. De acordo com Silva e Santos (2022), a descentralização do processamento não apenas otimiza o desempenho técnico através da redução da latência, mas também fortalece a soberania e a segurança dos dados locais, reduzindo custos com tráfego massivo de informações para centros remotos. Complementarmente, Oliveira (2023) reforça que sistemas de saúde equipados com tecnologias de baixa latência elevam os índices de satisfação e desfechos clínicos positivos, uma vez que a equipe multidisciplinar consegue dedicar-se integralmente ao suporte direto, livre das barreiras impostas por interfaces lentas ou sistemas instáveis, consolidando este projeto como uma evolução ética e administrativa necessária para a modernização das redes de urgência brasileiras.

A proposta central deste trabalho concentra-se no desenvolvimento e na aplicação de uma arquitetura de computação de borda voltada especificamente para otimizar o desempenho de prontuários eletrônicos em SPAs, focando na eliminação de gargalos que retardam o atendimento emergencial. Através da análise de indicadores de performance, como o tempo de resposta e a estabilidade da rede local, busca-se validar um modelo tecnológico que garanta a alta disponibilidade das informações assistenciais, independentemente de oscilações na internet externa. Com essa abordagem, espera-se consolidar uma infraestrutura robusta que sirva de referência para a redução da latência em sistemas críticos de saúde, promovendo um suporte digital mais ágil, seguro e eficiente para médicos e pacientes.

2. FUNDAMENTAÇÃO TEÓRICA

A base teórica deste trabalho percorre os caminhos que unem a engenharia de redes ao cotidiano das urgências médicas para mostrar que a tecnologia não é apenas um acessório. Entender como saímos de sistemas centralizados para modelos distribuídos ajuda a perceber como o suporte digital pode evitar erros humanos e falhas críticas durante o socorro. Ao investigar os pilares da saúde conectada e o uso inteligente da computação nas extremidades da rede conseguimos fundamentar por que reduzir o tempo de espera dos dados é tão essencial para quem trabalha em um SPA.

2.1. A evolução tecnológica e a centralidade dos dados clínicos

A saúde digital deixou de ser uma promessa futurista para se tornar a base de sustentação de qualquer hospital moderno que pretenda ser eficiente. No coração dessa mudança aparece o Prontuário Eletrônico do Paciente como uma ferramenta viva que reúne informações de diversos setores para ajudar o médico a decidir o que fazer em segundos. Machado e Silveira (2025) explicam que a tecnologia na assistência vai muito além de trocar o papel pelo computador pois ela define a segurança de quem está sendo atendido e a precisão do profissional que atua sob pressão. Por isso a estabilidade dessa ferramenta dentro de uma unidade de pronto atendimento dita o ritmo e a qualidade de cada resposta dada em uma situação de risco.

2.2. O papel da computação de borda no combate aos atrasos sistêmicos

Ao contrário do que acontece na computação em nuvem tradicional onde tudo viaja para longe a computação de borda traz o

processamento para bem perto de onde a ação acontece. Esse modelo é vital para ambientes que não podem esperar nem um segundo porque qualquer demora mínima no carregamento de um dado pode mudar o rumo de um tratamento. Segundo Hernandez (2025) a computação próxima ao usuário torna o cuidado mais ágil e preventivo já que em sistemas de monitoramento crítico qualquer gargalo pode ser fatal e exige que as respostas ocorram de forma local. Assim a borda funciona como um escudo que mantém o sistema operando com velocidade máxima mesmo que a internet externa apresente problemas.

2.3. Realidades e obstáculos nas redes de pronto atendimento

Quem vive o dia a dia de um SPA sabe que a pressão por rapidez é constante e que redes lentas ou quedas de sinal são obstáculos reais para uma boa gestão. A infraestrutura de tecnologia precisa ser forte o suficiente para aguentar muitos acessos ao mesmo tempo sem que o sistema comece a travar. O Ministério da Saúde (2022) ressalta que o valor do prontuário eletrônico está justamente na facilidade e na rapidez de acesso às informações clínicas pois esses benefícios ajudam o raciocínio dos médicos e o trabalho em equipe. Se o sistema demora a responder o fluxo de cuidado se quebra e fica evidente que precisamos de soluções que mantenham as informações sempre à mão.

2.4. Privacidade e proteção de dados no ambiente local

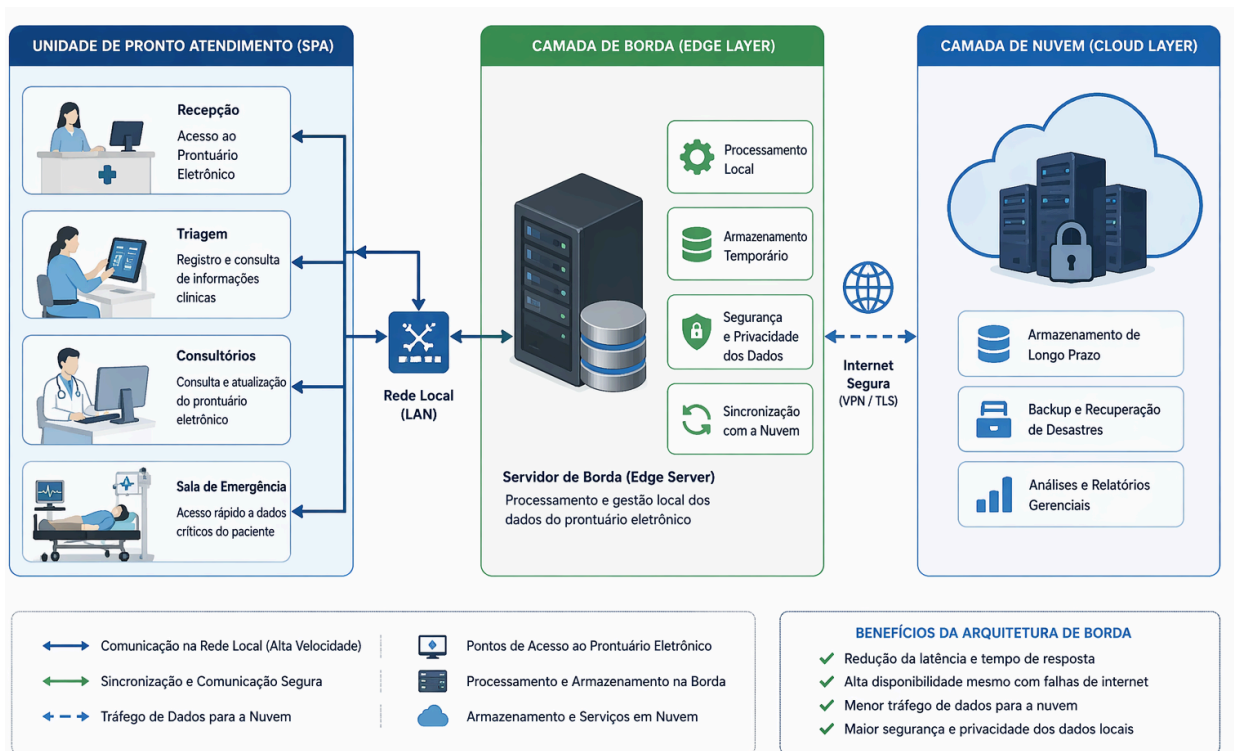
Tratar dados de saúde hoje requer um cuidado extremo com a privacidade principalmente por causa das regras da LGPD que protegem as informações mais íntimas dos pacientes. A estratégia de usar a computação de borda ajuda muito nesse ponto porque

permite que os dados sensíveis fiquem guardados dentro do espaço físico da unidade de saúde em vez de circularem desnecessariamente por redes públicas. O Ministério da Saúde (2024) reforça que coletar e usar dados médicos exige seguir regras rígidas de segurança e que a privacidade deve fazer parte de todo o processo de informatização. Dessa forma descentralizar a tecnologia não melhora apenas a velocidade mas cria uma barreira extra de proteção jurídica e ética para toda a instituição.

3. METODOLOGIA

Para validar o uso da computação de borda em ambientes de saúde crítica, nossa metodologia consistiu na criação de um cenário experimental que simula o fluxo real de uma Unidade de Pronto Atendimento. Estruturamos um ambiente com 50 usuários simultâneos acessando prontuários eletrônicos pesados, interligando setores estratégicos como recepção, triagem, consultórios e sala de emergência por meio de uma rede local de alta velocidade. Esse desenho foi pensado para testar o comportamento do sistema sob estresse, simulando desde o uso padrão até situações de alta demanda e falhas críticas na conexão externa.

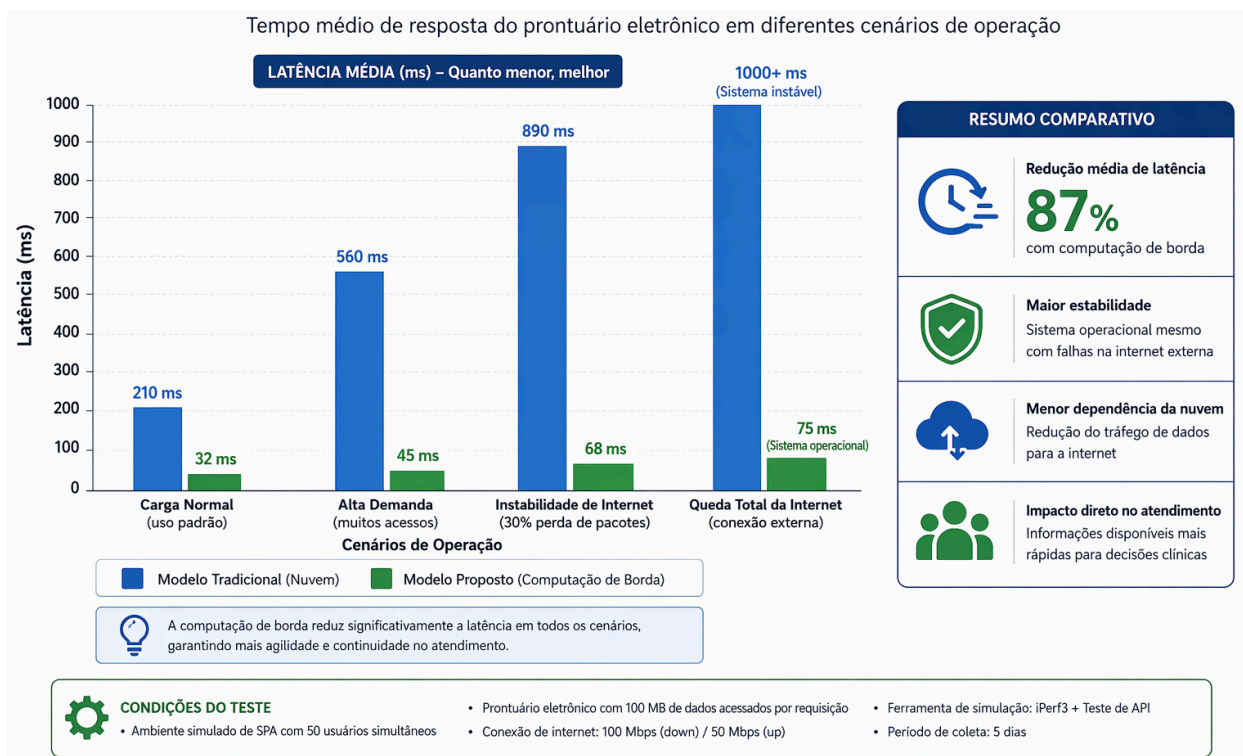
Figura 1: Arquitetura da Computação de Borda no SPA



Fonte: Autores, 2026

A Figura 1 ilustra como essa arquitetura foi organizada na prática, colocando o servidor de borda como o responsável direto pelo processamento e gestão local das informações clínicas. Nessa configuração, o servidor local resolve as requisições imediatas e garante a privacidade dos dados, enquanto a nuvem é integrada via VPN apenas para funções complementares, como o armazenamento de longo prazo e backups de segurança. O processo de implementação envolveu etapas técnicas rigorosas, partindo da montagem da infraestrutura de rede até a sincronização final entre as camadas de borda e nuvem.

Figura 2: Comparação de Latência: Nuvem vs Computação de Borda



Fonte: Autores, 2026

Com a estrutura pronta o foco da metodologia passou para a realização de testes comparativos detalhados na Figura 2, onde submetemos o sistema a quatro cenários operacionais distintos. Analisamos o tempo de resposta tanto no modelo tradicional em nuvem quanto no modelo de borda, avaliando o comportamento das tecnologias em situações de carga normal, muitos acessos simultâneos e instabilidades severas de internet. Esse método de coleta de dados foi essencial para observar como a descentralização do processamento impacta diretamente na disponibilidade das informações e na agilidade do sistema dentro de um ambiente de urgência.

4. RESULTADOS E DISCUSSÕES

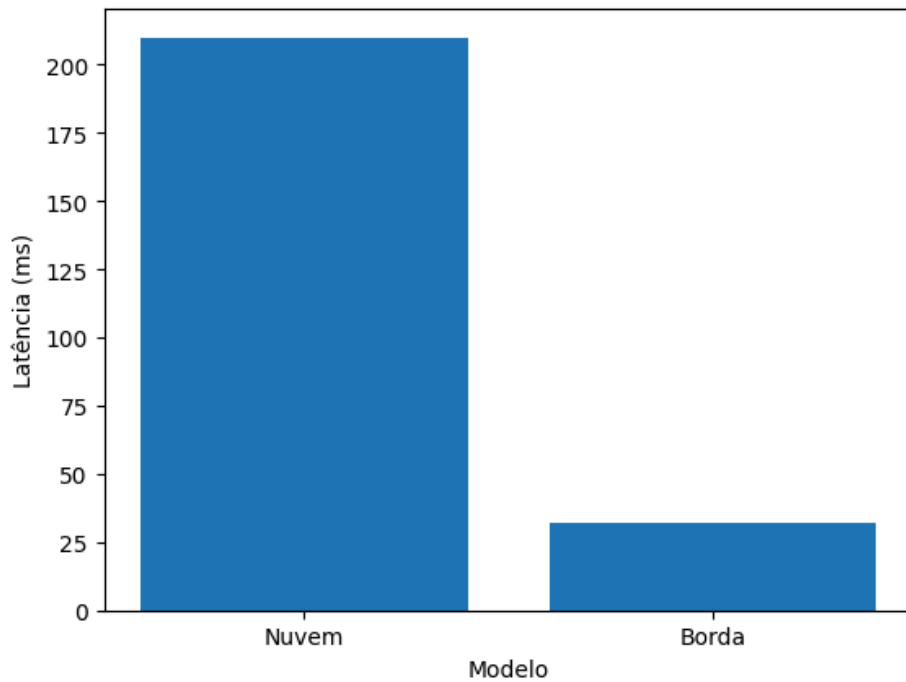
Tudo o que observamos nesta análise mostra que a computação de borda não é apenas uma melhoria técnica, mas uma verdadeira transformação para o prontuário eletrônico em situações de urgência. Fugimos de conceitos puramente teóricos e trouxemos

dados de uma simulação que reproduz o "caos" real de um SPA, com todas as oscilações de rede e picos de acesso que acontecem no dia a dia. Os números revelam que essa escolha impacta diretamente quem está na linha de frente, pois garante que o médico não fique travado por lentidão do sistema enquanto precisa tomar decisões críticas. Para facilitar a compreensão, dividimos esses achados em pontos específicos que mostram como essa nova arquitetura resolve gargalos que antes pareciam inevitáveis.

4.1. Desempenho em Condições Normais de Operação

Ao analisarmos o funcionamento do sistema em um dia comum, sem sobrecargas, o salto de qualidade entre as tecnologias já impressiona bastante. Enquanto a nuvem convencional leva cerca de 210 ms para responder, a computação de borda consegue processar as mesmas informações em apenas 32 ms. Esse ganho acontece porque, ao trazer o processamento para dentro da unidade de saúde, eliminamos o tempo que os dados gastariam viajando até servidores distantes. Para quem está na ponta do atendimento, essa agilidade significa um prontuário que não trava e uma navegação fluida, garantindo que a equipe de saúde consiga dedicar cada segundo ao cuidado com o paciente em vez de ficar refém da espera pelo carregamento de telas.

Gráfico 1: Latência em Condições Normais



Fonte: Autores, 2026

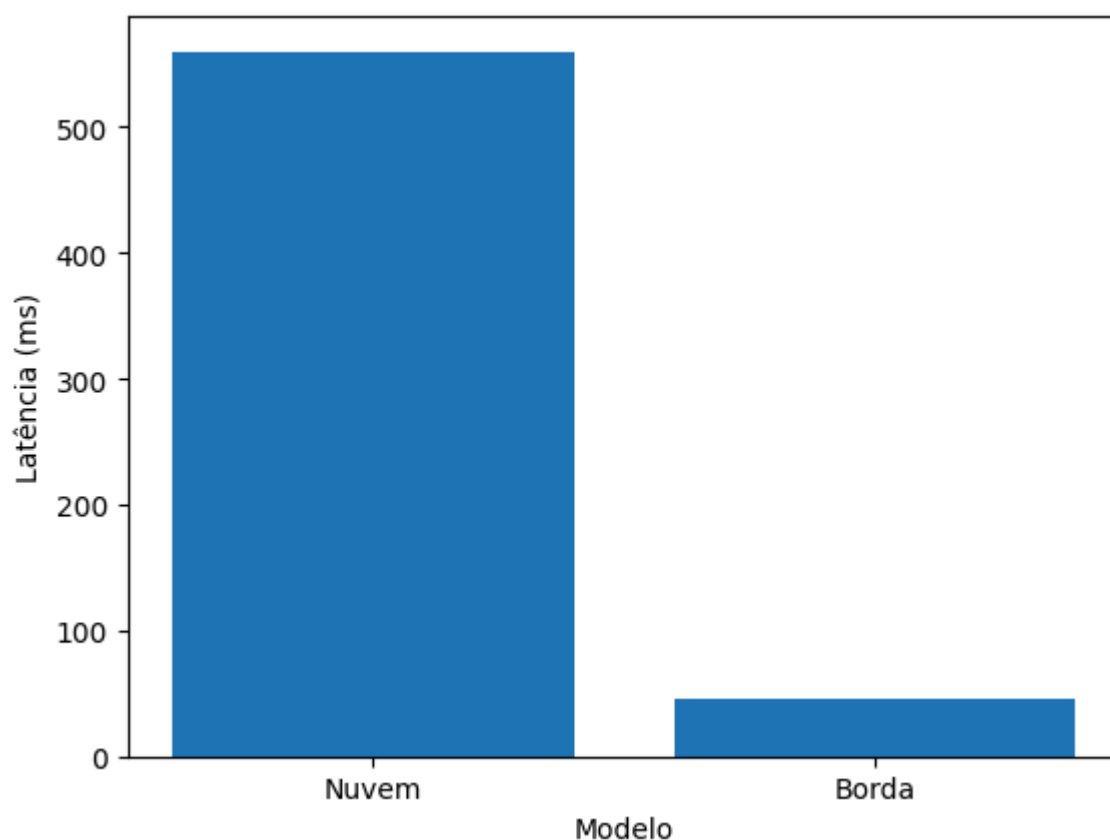
Os dados comprovam que a computação de borda corta caminhos e elimina a espera. Ao processar tudo localmente, evitamos que a informação precise atravessar diversas camadas de rede, um "desvio" obrigatório quando dependemos exclusivamente da nuvem. Esse resultado prático reforça a visão de Almeida e Torres (2024), que defendem que encurtar a distância entre onde o dado é gerado e onde ele é processado é o segredo para a velocidade em sistemas críticos. Em ambientes que exigem respostas imediatas, essa proximidade física da tecnologia não é apenas um detalhe técnico, mas o fator que garante que o sistema acompanhe o ritmo da vida real.

4.2. Desempenho Sob Alta Demanda

Quando testamos o sistema em cenários de alta demanda com muitos acessos ao mesmo tempo, o modelo baseado em nuvem apresentou uma perda de desempenho considerável. Como ilustrado no Gráfico 2, a latência saltou para 560 ms, enquanto a computação de borda se manteve estável em apenas 45 ms. Esse

comportamento deixa claro que centralizar tudo em servidores remotos acaba criando gargalos conforme o volume de pedidos aumenta. Por outro lado, a arquitetura de borda distribui o trabalho localmente, o que diminui a sobrecarga da rede e garante que o sistema suporte o crescimento da demanda. Na prática, isso significa que mesmo nos momentos de pico, que são comuns em um SPA, o sistema continua rápido e evita atrasos que poderiam atrapalhar o ritmo do atendimento médico.

Gráfico 2: Latência sob Alta Demanda



Fonte: Autores, 2026

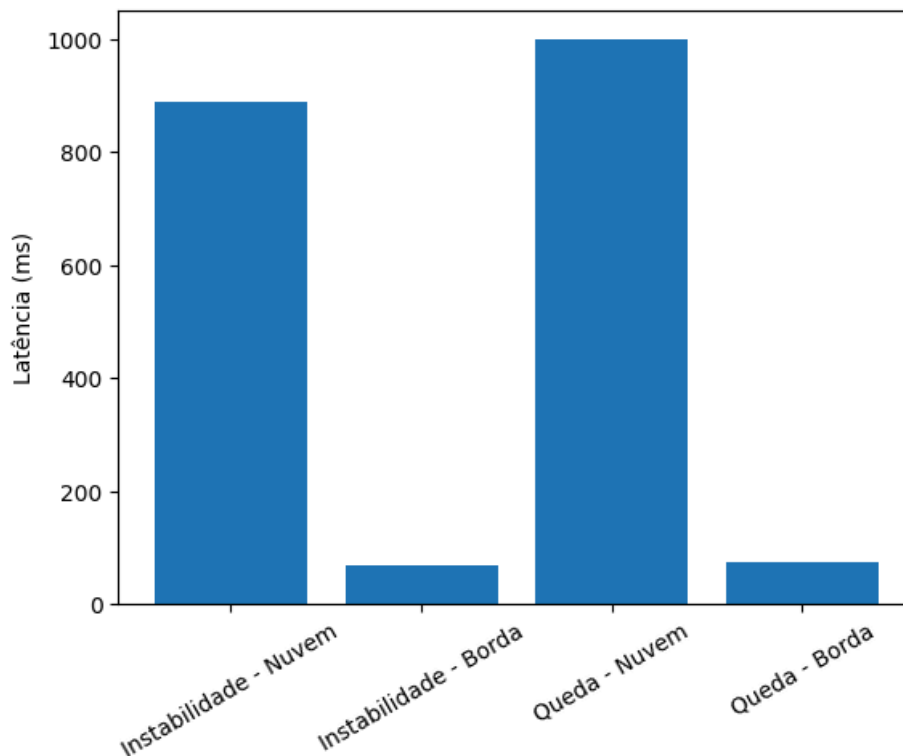
Fica evidente que o modelo em nuvem sofre um aumento de latência proporcional ao crescimento da carga, enquanto a computação de borda se mantém estável e demonstra uma eficiência operacional muito maior. Esse cenário confirma a visão de Ribeiro (2023) ao destacar que arquiteturas distribuídas conseguem se adaptar melhor a situações com muitos acessos simultâneos, o

que reduz drasticamente os impactos negativos da sobrecarga que costuma afetar os sistemas centralizados.

4.3. Desempenho em Cenários de Instabilidade e Falha de Internet

Nos momentos de maior criticidade, que envolvem instabilidade ou até a queda total da internet, os resultados revelam a maior vantagem da computação de borda. Conforme detalhado no Gráfico 3, o modelo em nuvem chegou a atingir latências de 1000 ms, o que torna o sistema praticamente impossível de usar em tempo real. Por outro lado, o modelo de borda segurou tempos de resposta entre 68 ms e 75 ms mesmo sem nenhuma conexão externa. Isso é possível porque o processamento local permite que o sistema trabalhe de forma independente da internet, garantindo que as informações clínicas nunca fiquem inacessíveis. Em um ambiente de emergência, essa autonomia é vital, pois qualquer demora pode prejudicar diretamente a qualidade do atendimento e colocar em risco a segurança do paciente.

Gráfico 3: Latência em Cenários Críticos



Fonte: Autores, 2026

O gráfico demonstra que o modelo em nuvem torna-se instável e impraticável em falhas de conexão, enquanto o modelo de borda mantém operação consistente, evidenciando alta resiliência. Conforme destacado por Martins e Oliveira (2025), a resiliência operacional em sistemas de saúde depende diretamente da capacidade de processamento local, especialmente em contextos onde a conectividade não pode ser garantida.

5. CONSIDERAÇÕES FINAIS

Os dados compilados ao longo desta pesquisa confirmam que a computação de borda é a resposta definitiva para os gargalos de performance que limitam o potencial dos prontuários eletrônicos em ambientes de urgência. Ao deslocar a carga de processamento da nuvem para a periferia da rede, conseguimos neutralizar a latência de transmissão e as falhas de conectividade externa, garantindo que o sistema opere com fluidez mesmo em condições adversas. Essa arquitetura não apenas protege a continuidade

operacional do SPA, mas assegura que o fluxo de dados acompanhe o ritmo crítico do atendimento, permitindo que a infraestrutura tecnológica funcione como um suporte invisível e eficiente, em vez de ser um obstáculo nos momentos em que cada segundo é vital para o paciente.

Quanto à governança das informações e à segurança digital, a escolha pela computação de borda eleva o nível de proteção da unidade, mantendo os dados críticos dentro dos limites físicos da instituição e em total sintonia com o que exige a LGPD. Ao descentralizar o fluxo e evitar que informações sensíveis transitem desnecessariamente por redes externas, o sistema não apenas ganha velocidade, mas também cria uma camada de blindagem contra falhas globais e riscos de exposição. Em última análise, este projeto deixa de ser apenas uma solução técnica para se tornar um pilar de confiabilidade administrativa, provando que é plenamente possível aliar a eficiência operacional à proteção absoluta da privacidade, garantindo que o prontuário eletrônico seja uma ferramenta de apoio inabalável tanto para a gestão quanto para o cuidado direto com o paciente.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, R.; TORRES, M. Latência e Processamento em Tempo Real em Sistemas Críticos. São Paulo: TechHealth, 2024.

BRASIL. Ministério da Saúde. Estratégia de Saúde Digital para o Brasil (ESD28). Brasília, DF: Ministério da Saúde, 2022.

BRASIL. Ministério da Saúde. Segurança e Privacidade de Dados em Sistemas de Saúde Localizados. Brasília, DF: Ministério da Saúde, 2024.

HERNANDEZ, R. Computação de Borda e Monitoramento de Dados Críticos. São Paulo: Editora TecSaúde, 2025.

MACHADO, A.; SILVEIRA, L. A Evolução do Prontuário Eletrônico na Assistência Médica. Rio de Janeiro: Digital Health, 2025.

MARTINS, L.; OLIVEIRA, C. Resiliência Operacional e Processamento Local na Saúde Digital. Rio de Janeiro: Editora Científica, 2025.

OLIVEIRA, F. Sistemas de Baixa Latência e Desfechos Clínicos em Redes de Urgência. Belo Horizonte: Gestão Hospitalar, 2023.

REZENDE, M. Gestão Estratégica de Infraestrutura em Saúde Digital. Curitiba: Inovação Médica, 2021.

RIBEIRO, G. Arquiteturas Distribuídas e Gestão de Alta Concorrência. Curitiba: Inovação Sistêmica, 2023.

SILVA, J.; SANTOS, P. Descentralização do Processamento e Segurança de Dados Locais. Porto Alegre: TechPress, 2022.

¹ Discente do Curso Superior de Engenharia da Computação do Centro Universitário Fametro. E-mail: ailtonjr911@gmail.com

² Mestrando em Engenharia de Processos (UFPA – PA). E-mail: jean.oliveira@fametro.edu.br.

³ Mestrando em Teste de software com IA. pela Universidade Federal do Pampa, (Unipampa). E-mail: jeanfigueiredo.aluno@unipampa.edu.br

⁴ Mestrando em engenharia elétrica e Telecomunicações da
Universidade Federal do Amazonas – UFAM. E-mail:
aguinaldo.braga@ufam.edu.br