

# Evoluindo a Maturidade FinOps: A Jornada do Inter rumo à Automação Inteligente de Kubernetes



## O Inter

Fundado na década de 1980, o Inter consolidou-se como um dos principais players de tecnologia financeira do mundo. Listado na Nasdaq e pioneiro como o primeiro banco digital 100% gratuito do Brasil, o Inter opera hoje um Super App que atende aproximadamente 40 milhões de clientes. Com a meta estratégica de atingir 60 milhões de usuários até 2027, a instituição demanda uma infraestrutura tecnológica de alta performance, capaz de sustentar um crescimento exponencial com governança e eficiência técnica.



## A EVOLUÇÃO PARA A GESTÃO DINÂMICA DE RECURSOS

A parceria estratégica entre o Inter e a RealCloud já colhe frutos consolidados desde sua primeira fase em 2024, a partir do foco central na adoção da prática de Spot Optimization. O sucesso dessa prática ajudou o Inter a avançar ainda mais na maturidade em FinOps, atingindo o nível de 70% de utilização de instâncias Spot da sua capacidade computacional. No entanto, o rápido amadurecimento dos ambientes cloud-native trouxe novos desafios: à medida que a infraestrutura se tornava mais dinâmica e complexa. A manutenção da eficiência passou a exigir uma atenção manual constante das equipes de engenharia, que precisavam "perseguir" o ajuste ideal de recursos de forma ininterrupta.

*"Alcançar 70% de economia em compute com automação de instâncias spot foi um marco, mas a verdadeira maturidade FinOps vem ao eliminar o ajuste fino manual. Automatizar o rightsizing transformou nossa eficiência de uma tarefa reativa em um pilar nativo da nossa governança de nuvem."*

**Victoria Ribeiro**  
FinOps Head



O cenário revelou um gap crítico na gestão de infraestrutura: enquanto a alocação de recursos permanecia estática, o consumo e a demanda no mundo real eram inerentemente dinâmicos. No Inter, o processo de rightsizing — o ajuste fino na alocação de CPU e memória para cada workload — demandava uma velocidade incompatível com processos não automatizados. Engenheiros dedicavam um tempo precioso ajustando recursos para acompanhar as flutuações de demanda, e ainda assim, frequentemente isso resultava em desperdício de capital com recursos ociosos ou riscos à performance durante picos de acesso. Ficou evidente que, para sustentar a escala de um Super App, a melhor solução seria a automação total dos recursos de nuvem em tempo de execução (run-time).

## ESTRATÉGIA E SINERGIA TECNOLÓGICA

Como parceira estratégica e implementadora global de FinOps, a RealCloud atuou como advisor para desenhar o próximo salto de maturidade do Inter, apresentando uma solução de Kubernetes Pod Rightsizing. A solução foi ancorada na plataforma ScaleOps como a tecnologia capaz de gerenciar essa complexidade de forma autônoma. A implementação focou no impacto que o dimensionamento de Pods causa na alocação de Nodes, criando um ciclo de eficiência exponencial. Ao automatizar o gerenciamento em tempo de execução, a solução passou a identificar e executar o ajuste fino dos requests dos contêineres instantaneamente, eliminando a latência das decisões humanas.

"A integração do ScaleOps permitiu que nossos clusters Kubernetes realizassem ajustes de requests em tempo real, sem latência humana. Agora, cada microsserviço consome exatamente o necessário, garantindo performance máxima com desperdício zero."

**Matheus Meneses**  
Principal Platform Engineer



Essa orquestração garantiu que a infraestrutura reagisse em tempo real à liberação de recursos, executando o bin-packing e o escalonamento para baixo das instâncias de forma precisa. A transição para o modelo autônomo permitiu que o Inter fosse além da simples observabilidade, implementando uma inteligência que garante que cada micro serviço utilize exatamente o necessário para sua performance. Essa abordagem removeu o custo de oportunidade da equipe de engenharia, permitindo que o foco fosse redirecionado para a inovação, enquanto a infraestrutura operava de forma autossuficiente e otimizada.

## RESULTADOS E IMPACTO NO NEGÓCIO

A automação do rightsizing entregou resultados tangíveis que reforçam a sustentabilidade do modelo de negócio do Inter. Em ambientes produtivos, a parceria alcançou uma redução adicional de 33% nos custos de computação,

gerando uma economia anual projetada superior a US\$ 1 milhão. Além do impacto financeiro, a eficiência técnica foi notável, com a otimização da densidade de clusters reduzindo o volume total de instâncias, antes na faixa de 3.000 para uma média de 2.500, operando com maior inteligência de recursos e sem qualquer prejuízo à experiência do usuário final.

"Reduzir 33% dos custos e otimizar 500 instâncias sem afetar a experiência do usuário prova que a automação é o caminho para a rentabilidade. Essa economia anual superior a US\$ 1 milhão reafirma nossa vanguarda tecnológica e financeira."

**Carlos Pedrosa**  
IT Director



O sucesso desta etapa solidifica a RealCloud como um braço de consultoria contínua para o Inter, com um roadmap que contempla otimizações em outras camadas além da de Compute e estratégias avançadas de posicionamento de cargas de trabalho. A parceria prova que a maturidade em nuvem não é um destino estático, mas um processo de inovação constante. Ao automatizar a eficiência, o Inter assegura que sua tecnologia seja um catalisador de rentabilidade, mantendo a vanguarda no setor financeiro global.



+55 (21) 2179 7788

### Sobre a RealCloud

A RealCloud é líder em soluções de gestão e otimização de nuvem na América Latina, ajudando empresas a alcançar eficiência operacional com redução de custos e aumento de performance. Com clientes como VTEX, iFood e Hotmart, a RealCloud já auxiliou na economia de mais de US\$ 440 milhões em gastos com nuvem. Saiba mais em: [realcloud.ai](https://realcloud.ai)



**RealCloud**

© 2026 por RealCloud