

Evolucionando la Madurez FinOps: El Viaje de Inter hacia la Automatización Inteligente de Kubernetes



El Inter

Fundado en la década de 1980, Inter se ha consolidado como uno de los principales players de tecnología financiera del mundo. Listado en el Nasdaq y pionero como el primer banco digital 100% gratuito de Brasil, Inter opera hoy una Super App que atiende aproximadamente a 40 millones de clientes. Con la meta estratégica de alcanzar los 60 millones de usuarios para 2027, la institución demanda una infraestructura tecnológica de alto rendimiento, capaz de sostener un crecimiento exponencial con gobernanza y eficiencia técnica.



LA EVOLUCIÓN HACIA LA GESTIÓN DINÁMICA DE RECURSOS

La alianza estratégica entre Inter y RealCloud ya cosecha frutos consolidados desde su primera fase en 2024, a partir del enfoque central en la adopción de la práctica de Spot Optimization. El éxito de esta práctica ayudó a Inter a avanzar aún más en la madurez de FinOps, alcanzando el nivel de 70% de utilización de instancias Spot de su capacidad computacional. Sin embargo, la rápida maduración de los entornos cloud-native trajo nuevos desafíos a medida que la infraestructura se volvía más dinámica y compleja. El mantenimiento de la eficiencia pasó a exigir una atención manual constante de los equipos de ingeniería, que necesitaban "perseguir" el ajuste ideal de recursos de forma ininterrumpida.

"Alcanzar un 70% de ahorro en compute con la automatización de instancias spot fue un hito, pero la verdadera madurez FinOps surge al eliminar el ajuste fino manual. Automatizar el rightsizing transformó nuestra eficiencia de una tarea reactiva en un pilar nativo de nuestra gobernanza de nube".

Victoria Ribeiro
FinOps Head



El escenario reveló una brecha crítica en la gestión de infraestructura: mientras la asignación de recursos permanecía estática, el consumo y la demanda en el mundo real eran inherentemente dinámicos. En Inter, el proceso de rightsizing —el ajuste fino en la asignación de CPU y memoria para cada workload— demandaba una velocidad incompatible con procesos no automatizados. Los ingenieros dedicaban un tiempo precioso ajustando recursos para acompañar las fluctuaciones de demanda y, aun así, esto frecuentemente resultaba en desperdicio de capital con recursos ociosos o riesgos para el rendimiento durante picos de acceso. Quedó evidente que, para sostener la escala de una Super App, la mejor solución sería la automatización total de los recursos de nube en tiempo de ejecución (run-time).

ESTRATEGIA Y SINERGIA TECNOLÓGICA

Como socio estratégico e implementador global de FinOps, RealCloud actuó como asesor para diseñar el próximo salto de madurez de Inter, presentando una solución de Kubernetes Pod Rightsizing. La solución se ancló en la plataforma ScaleOps como la tecnología capaz de gestionar esta complejidad de forma autónoma. La implementación se centró en el impacto que el dimensionamiento de Pods causa en la asignación de Nodes, creando un ciclo de eficiencia exponencial. Al automatizar la gestión en tiempo de ejecución, la solución pasó a identificar y ejecutar el ajuste fino de los requests de los contenedores instantáneamente, eliminando la latencia de las decisiones humanas.

"La integración de ScaleOps permitió que nuestros clústeres de Kubernetes realizaran ajustes de requests en tiempo real, sin latencia humana. Ahora, cada microservicio consume exactamente lo necesario, garantizando el máximo rendimiento con desperdicio cero".

Matheus Meneses
Principal Platform Engineer



Esta orquestación garantizó que la infraestructura reaccionara en tiempo real a la liberación de recursos, ejecutando el bin-packing y el escalado hacia abajo de las instancias de forma precisa. La transición al modelo autónomo permitió que Inter fuera más allá de la simple observabilidad, implementando una inteligencia que garantiza que cada microservicio utilice exactamente lo necesario para su rendimiento. Este enfoque eliminó el costo de oportunidad del equipo de ingeniería, permitiendo que el foco se redirigiera hacia la innovación, mientras la infraestructura operaba de forma autosuficiente y optimizada.

RESULTADOS E IMPACTO EN EL NEGOCIO

La automatización del rightsizing entregó resultados tangibles que refuerzan la sostenibilidad del modelo de negocio de Inter. En entornos productivos, la alianza alcanzó una reducción adicional del 33% en los costos de computación,

generando un ahorro anual proyectado superior a US\$ 1 millón. Además del impacto financiero, la eficiencia técnica fue notable, con la optimización de la densidad de clústeres reduciendo el volumen total de instancias, antes en el rango de 3,000 para un promedio de 2,500, operando con mayor inteligencia de recursos y sin perjuicio alguno para la experiencia del usuario final.

"Reducir el 33% de los costos y optimizar 500 instancias sin afectar la experiencia del usuario demuestra que la automatización es el camino hacia la rentabilidad. Este ahorro anual superior a US\$ 1 millón reafirma nuestra vanguardia tecnológica financiera".

Carlos Pedrosa
IT Director



El éxito de esta etapa consolida a RealCloud como un brazo de consultoría continua para Inter, con un roadmap que contempla optimizaciones en otras capas además de la de Compute y estrategias avanzadas de posicionamiento de cargas de trabajo. La alianza demuestra que la madurez en la nube no es un destino estático, sino un proceso de innovación constante. Al automatizar la eficiencia, Inter asegura que su tecnología sea un catalizador de rentabilidad, manteniendo la vanguardia en el sector financiero global.



+55 (21) 2179 7788

Sobre RealCloud

RealCloud es líder en soluciones de gestión y optimización de nube en América Latina, ayudando a las empresas a alcanzar la eficiencia operativa con reducción de costos y aumento de rendimiento. Con clientes como VTEX, iFood y Hotmart, RealCloud ya ha ayudado a ahorrar más de US\$ 440 millones en gastos de nube. Más información en: realcloud.ai



RealCloud

© 2026 por RealCloud