



Company: iFood
Industry: E-commerce
Location: São Paulo, Brazil

“Nossos modelos ML oferecem múltiplas recomendações durante uma única sessão de usuário. Precisamos de um banco de dados de alto desempenho para lidar com estas operações de leitura e escrita. Redis Enterprise Cloud resolve este problema muito bem para nós”.

–Daniel Galinkin, Head de Machine Learning Engineering, iFood

O Cliente

Os serviços de encomenda e entrega de alimentos on-line do iFood são amplamente utilizados em todo o Brasil e na Colômbia. Com sede em São Paulo, iFood mantém uma participação de 80% no mercado brasileiro de entrega de alimentos.

A solução

Para garantir a consistência dos engenheiros ML e criar uma experiência ótima para os usuários, o iFood usa Redis Enterprise Cloud na AWS na fundação de sua loja de recursos ML em rápida evolução. O Redis on Flash maximiza o rendimento do processamento de dados enquanto reduz os custos gerais de armazenamento de dados.

O Desafio

Para as aplicações populares de pedidos e entregas de alimentos do iFood, o sucesso é um subproduto do desempenho: os modelos de aprendizagem de máquinas (ML) do iFood devem ser capazes de processar dados rapidamente para reduzir custos, aumentar a receita e influenciar positivamente o comportamento do usuário durante as interações em tempo real.

Os benefícios

A infra-estrutura baseada no iFood Redis oferece um desempenho de sub-milissegundo para operações de ML, alimentando uma arquitetura de microserviços altamente eficiente que otimiza a experiência do usuário, permite uma escalabilidade maciça e permite um rápido crescimento corporativo.

A pandemia COVID-19 apresentou oportunidades únicas para empresas de comércio eletrônico, especialmente serviços de entrega on-line que estavam preparados para lidar com um volume crescente de pedidos. No iFood, uma das maiores empresas de entrega de alimentos da América Latina, a equipe de tecnologia se transformou em uma engrenagem importante à medida que milhões de novos usuários começaram a confiar em seu serviço, e milhares de novos restaurantes se juntaram à sua crescente plataforma. Somente nos primeiros três meses de 2021, mais de 30.000 novos restaurantes, mercearias e lojas de conveniência se inscreveram para entregar alimentos através da plataforma digital do iFood, e seus pedidos mensais aumentaram em dezenas de milhões de entregas.¹

Apesar deste aumento no volume de negócios e da crescente receita que o acompanhou, iFood manteve-se fiel aos valores que impulsionaram a empresa em rápido crescimento para o sucesso. “Não se trata apenas de fazer um pedido de comida”, resume Daniel Galinkin, chefe de inteligência artificial e aprendizagem de máquinas no iFood. “Trata-se de proporcionar uma experiência ótima para nossos clientes”. Essa é uma das coisas que distingue nossa empresa, e faz com que as pessoas voltem”.

Serviço de Concierge: Usando a IA para otimizar a experiência do usuário

Hoje, o iFood atende uma rede de cerca de 320.000 restaurantes e tem como objetivo aumentar sua base de 30.000 supermercados.² Recentemente, também acrescentou a entrega para farmácias e lojas de bebidas. Para cumprir estes elevados objetivos de crescimento, Daniel Galinkin e os outros membros da

equipe de Data&AI do iFood utilizam a inteligência artificial (IA) e o aprendizado de máquinas para entender melhor os usuários, tais como rastrear quantos pedidos fizeram no último mês, quais restaurantes e lojas preferem, seus mecanismos de pagamento escolhidos e muitas outras variáveis. Estes fatos são mantidos como características - atributos e variáveis únicas que servem como inputs para os modelos ML do iFood.

Para garantir uma experiência ótima do usuário, o iFood deve ser capaz de executar esses modelos à medida que os pedidos são feitos para influenciar as decisões de segundo grau que os clientes tomam. “Uma única característica pode incorporar várias regras comerciais, cada uma das quais está em constante evolução”, explica Galinkin. “Toda essa lógica e todas essas definições devem ser consistentemente compartilhadas entre as aplicações, ou um usuário pode ser tratado de forma diferente em diferentes lugares no aplicativo. Se você não centralizar este conhecimento em uma pontuação de recurso, os esforços da equipe se tornam fragmentados”.

O iFood tem mais de 100 modelos ML em produção. Alguns desses modelos ajudam a trazer receita, como os que alimentam seu motor de recomendação (que ajuda os usuários a decidir quais restaurantes devem ser considerados e quais alimentos pedir). Outros modelos determinam que tipos de vales e cupons devem ser apresentados aos usuários em cada momento da navegação. Ainda outros modelos ajudam iFood a reduzir custos detectando transações fraudulentas, tais como chargebacks, em que os usuários fazem pedidos, recebem esses pedidos, e depois tentam receber um reembolso em seus cartões de crédito.



¹ “A tecnologia alimentar brasileira iFood, o gigante iFood, entrega 60 milhões de pedidos mensais em meio à pandemia”. *Latin American Business Stories (LABS)*, 12 de abril de 2021

² “Brazil Delivery App iFood Focusing on Growth”, *U.S. News and World Report*, 1 de setembro de 2022

“Nosso negócio é bem sucedido quando conseguimos entregar os recursos de ML rapidamente”, diz Galinkin. “Nossos modelos ML entregam várias recomendações durante uma única sessão de usuário. Precisamos de um banco de dados de alto desempenho para lidar com essas operações de leitura e escrita”. Redis Enterprise Cloud resolve este problema muito bem para nós”.



Molho Secreto”: Redis Enterprise Cloud Incentiva Consistência para a Equipe MLOps

Uma loja de recursos é um sistema de gerenciamento de banco de dados que é otimizado para armazenar e servir dados de recursos em um banco de dados de valores-chave. As lojas de recursos permitem que os engenheiros reutilizem os recursos de ML ao longo da

descoberta de dados, treinamento, produção e outras fases do ciclo de vida da ciência dos dados.

Para garantir a consistência entre os desenvolvedores e criar uma experiência ótima para os usuários, o iFood utiliza o Redis Enterprise Cloud na AWS como base de sua loja de recursos em rápida evolução. “Um dos principais aspectos de ter feature store é o conhecimento compartilhado”, diz Galinkin. “O Redis Enterprise Cloud ajuda nossa equipe a colaborar”. Ele incentiva a reusabilidade para que não tenhamos que ‘inventar a roda’ várias vezes”.

Iniciativas bem sucedidas de aprendizado de máquina dependem da obtenção dos dados certos no momento certo para os modelos corretos. O [Redis Enterprise Cloud](#) disponibiliza dados de recursos a dezenas de modelos em produção, e inclui recursos essenciais como um registro, pipeline de dados e ferramentas de monitoramento para agilizar as atividades de engenharia de recursos. Isto permite que a equipe de Data&AI do iFood pesquise, reutilize e sirva recursos em produção, em escala. “Um armazenamento de recursos permite implementar o conhecimento uma vez e compartilhar os cálculos várias vezes”, acrescenta Galinkin.

Galinkin e sua equipe foram positivamente influenciados por seus colegas engenheiros de ML da DoorDash, outro serviço popular de entrega de alimentos que utiliza Redis para sua feature store. Eles gostaram especialmente da arquitetura aberta e baseada em padrões do ambiente Redis, que funciona bem com Databricks, Amazon SageMaker, Terraform, Apache Airflow, Python e outras ferramentas de software que ajudam a automatizar seus esforços.

Receita para o Sucesso: Redis on Flash Powers Características Carga de trabalho de engenharia

Em 2020, com o aumento dos negócios durante a pandemia de Covid, a equipe de ML Platform da iFood começou a construir uma feature store a partir do zero. Na época, não existiam soluções embaladas que se adequassem a eles, nem comerciais nem de código aberto. Inicialmente, eles gerenciavam sua feature store no ElastiCache da Amazon. Com o tempo, porém,, tornou-se muito caro colocar todos os seus dados de recursos em memória no AWS. “O ElastiCache era relativamente estável, mas realmente caro”, admite Galinkin. “Quando ouvimos falar da oferta Redis on Flash, decidimos dar uma olhada mais de perto”.



A Redis trabalhou de perto com o iFood para implementar o [Redis on Flash](#) e integrar o serviço totalmente gerenciado na arquitetura de microserviços do iFood. “Tivemos que migrar um enorme volume de dados, mas a transição foi um sucesso”, continua Galinkin. “Nunca fechamos um negócio e fizemos uma migração tão rápida”.

Hoje, sempre que um usuário inicia uma ação, como consultar um restaurante ou fazer um pedido, isso aciona vários micro-serviços. Segundo Galinkin, a latência combinada para todos esses microserviços não pode exceder 200 milissegundos. Cada milissegundo conta”, enfatiza ele, “por isso, realizamos alguns benchmarks”. A Amazon DynamoDB oferece cerca de 10 milissegundos por operação lida. O Redis Enterprise Cloud oferece menos de um milissegundo por operação de leitura. Isso faz diferença. Esse é um dos principais fatores que nos levou ao Redis Enterprise Cloud”.

A chave para este desempenho excepcional é a forma como a [Redis em Flash orquestra os dados entre as camadas DRAM e Flash Storage](#). Os dados mais acessados ou “quentes” residem na camada rápida de DRAM, juntamente com as chaves e o dicionário Redis. Os dados “frios” menos frequentemente acessados permanecem na camada local de armazenamento Flash. “O Redis on Flash faz com que seja perfeito para o usuário”, relata Galinkin. “Nós simplesmente emitimos pedidos e Redis entrega os dados”. É transparente para o cliente, e é muito, muito rápido - com desempenho de sub-milissegundo. Além disso, é muito mais barato [do que nossa solução anterior]”.

O iFood também foi influenciado pela excepcional estabilidade e confiabilidade do Redis Enterprise Cloud, que utiliza uma arquitetura de cluster sem compartilhamento para automatizar operações de failover em nível de processo, para nós individuais e entre zonas de disponibilidade de infra-estrutura. O iFood também utiliza o [Redis Enterprise Cloud como uma camada de cache](#) para armazenar os resultados das consultas, de modo que as solicitações repetidas de banco de dados possam ser atendidas instantaneamente.

“A tolerância a falhas é realmente importante para nossa aplicação de feature store, e o Redis Enterprise Cloud funciona muito bem como nosso banco de dados principal também”, conclui Galinkin. “Temos uma equipe pequena, portanto não queremos ficar atolados com a gestão da infra-estrutura”. O Redis Enterprise Cloud realmente nos ajuda com isto”. Isso nos poupa muito dinheiro em infra-estrutura”.

Saber mais sobre o Redis Enterprise como um [Feature Store](#).