

O papel estratégico dos *datalakes* na era da inteligência artificial: desafios e oportunidades

Por Danilo Zimmermann, diretor executivo de Tecnologia do Banco BS2

Segundo levantamento do IBGE, 85% das empresas com 100 ou mais colaboradores utilizaram pelo menos uma tecnologia digital avançada. Trabalho no mercado de tecnologia há mais de 20 anos e tenho testemunhado como cresce a busca por ferramentas que utilizam inteligência artificial.

Mas gostaria de dar um passo atrás e fazer uma reflexão sobre algo que permeia praticamente todas as tecnologias avançadas – os dados. Vitais para a competitividade das companhias e base para a transformação digital, eles possuem diversas maneiras de armazenamento, utilização, de interação e de uso. Visto que é uma matéria-prima, exige abordagens diferentes, como do ponto de vista de arquitetura e engenharia, da governança, da estratégia de uso e por aí vai.

Ressalto aqui a relevância da arquitetura e engenharia desses dados e como ela é imprescindível para uma jornada saudável de transformação digital de uma companhia. A questão é a seguinte: não adianta só ter onde colocar os dados, é preciso armazená-los de maneira organizada e eficiente. Na TI, o lugar onde os guardamos é chamado de *datalake*. Sem eles, não tem como avançar no uso de tecnologia avançada como é o caso de IA.

A maioria das empresas entendeu a importância de um bom *datalake*. Sabem também que é positivo processá-los em nuvem – segundo a mesma pesquisa de cima, 74% usam *cloud computing* em uma tentativa de deixar esses dados interoperáveis. E aí se vai muito dinheiro investido em repositórios em nuvem, capazes de armazenar de forma elástica bilhões de dados.

Porém, todo esse espaço na nuvem não tem muita valia se essas informações não estiverem organizadas de maneira que favoreça a sua empresa. É como ir ao mercado e demorar horas para achar um produto. Por isso, os *datalakes* demandam uma arquitetura/engenharia que facilitem e agilizem o acesso às informações, em um nível qualificado e em larga escala. O pulo do gato é ter não apenas um repositório abrangente, mas que seja também organizado e inteligente.

Vejo áreas de TI altamente oneradas pelo custo de manter esta estratégia ativa, sem um planejamento para deixá-lo de fato competitivo. O custo e o retorno baixo acabam desencorajando e as vezes exigindo voltar ao modelo de infraestrutura *on premises*. De acordo com uma previsão feita pela Índice de Desempenho de Custos (IDC) para 2024, os gastos com infraestrutura em nuvem serão de cerca de US\$ 129,9 bilhões, 19,3% a mais em comparação com 2023.

Por isso, um termo que tem sido bastante recorrente no mercado é o FinOps, que consiste em uma técnica de gerenciamento com o objetivo de promover responsabilidade financeira, principalmente no que diz respeito aos custos dos serviços na nuvem.

O uso dos dados não só para fazer relatórios, mas para gerar informações, insights que aumentem a assertividade na tomada de decisões deve ser o objetivo maior a ser buscado. É a partir desses *datalakes* estruturados que algoritmos imprescindíveis para o dia a dia dessas empresas são criados. E, mais uma vez, o papel do engenheiro de dados ganha importância neste contexto.

No caso dos bancos, os *datalakes* permitem por exemplo a construção de modelos inteligentes de concessão de crédito e de prevenção à lavagem de dinheiro (PLD). Neste caso, entram no jogo os cientistas de dados que, além do sistema, devem entender do negócio em si para criar modelos aderentes à indústria na qual ele está inserido.

Há pouco tempo, participei da 3ª edição do The North Summit, um evento que reúne lideranças de tecnologia de todo o mundo para discutir tendências e trocar experiências. A edição deste ano focou em inteligência artificial e IA Generativa, que utilizam amplamente os *datalakes* que mencionei acima. Mas, agora, com algoritmos, técnicas de capacidade de processamento aperfeiçoados.

Executivos de alto nível de diversos países trouxeram casos de suas empresas e é um consenso que a construção desses sistemas pode aumentar a produtividade de forma relevante nas áreas de TI e backoffice, reduzindo custos drasticamente. Porém, é fundamental ser estratégico nesta jornada cujo primeiro passo é acertar na arquitetura e engenharia de dados construindo *datalakes* performáticos e eficientes.