

Introdução a Arquitetura de Dados.

Para te apresentar o conceito de Arquitetura de Dados, vou contar uma sequência de atividades que fiz ontem. De manhã, acordei e fiz uma transferência para um fornecedor que trabalha comigo. Ao meio-dia, um entregador trouxe um produto que comprei pela internet. À noite, voltando para minha casa, pedi meu jantar através de um aplicativo. E, depois do jantar, fiz minha avaliação sobre a comida no mesmo app.

O que todos esses momentos têm em comum? Em todos eles, uma "transação" foi registrada. Isso significa que os dados gerados em cada transação precisam ser armazenados de forma segura e confiável em um banco de dados. Se você pensar sobre isso por alguns segundos, verá que muitas das suas ações diárias geram algum tipo de transação. Isso significa que, em apenas um dia, milhões de pessoas geram transações que, quando registradas nos aplicativos que usamos, geram milhões de dados.

Para que isso seja possível, há um sistema que chamamos OLTP, Online Transaction Processing ou processamento de transações online. OLTP é um tipo de processamento que facilita e gerencia as aplicações transacionais, onde usualmente as operações de entrada, recuperação e processamento de dados são executadas. E quais aplicações poderíamos dizer que são OLTP? São aplicações com as quais, por exemplo, uma empresa faz o registro de suas operações de negócios, suas interações com os seus clientes, fornecedores e funcionários, sua informação contábil, fiscal e legal, seus movimentos físicos de mercadorias e todas as aplicações web ou mobile, onde fazemos algum tipo de transação ou mesmo onde navegamos e fazemos consultas.

E por que estamos interessados nesse tipo de aplicação OLTP? Normalmente, na área de Data Analytics, os dados que serão consumidos e alimentarão nossos modelos analíticos são provenientes de aplicações OLTP. O principal objetivo do processamento transacional é registrar as transações de maneira eficiente e segura, além de otimizar o uso do espaço de armazenamento. Além do OLTP, há também o processamento OLAP, On-Line Analytical Processing, ou processamento analítico online, cujo objetivo é otimizar e tornar eficiente a análise de grandes volumes de dados.

Muitas vezes esse tipo de processamento é feito nas ferramentas de visualização, especialmente nas de última geração. Em quase todos os casos, as soluções OLAP incluem um recurso técnico que consiste em ter as informações pré-resumidas para que as consultas de dados que fazemos durante a análise sejam as mais rápidas possíveis. Para isso, é utilizado um modelo de design conceitual baseado em dois tipos de dados: dimensões e métricas.

As dimensões são aqueles dados ou campos do modelo que usamos para agrupar, filtrar, adicionar os dados que são medidos. Por exemplo: clientes, segmento de clientes, produto, família de produtos, ano, mês, faixa etária etc. Já as métricas são os valores nos quais podemos calcular as agregações, como somas, médias, máximos, mínimos etc. E agrupamos essas medidas pelas dimensões. Nos modelos de BI, por exemplo, geralmente são os nossos indicadores ou KPIs. Podemos citar como exemplo: vendas, custos, margem líquida, número de conversões por visita, tempo médio de entrega etc.

Em quase todas as ferramentas de Análise de Dados baseadas no modelo de processamento OLAP, encontramos dimensões e medidas ao definir nossos modelos de análise e nossas visualizações. Na maioria dos casos, é possível adicionar cálculos ou regras que nos permitem realizar operações entre métricas ou entre membros de uma dimensão, sem precisar de cálculos prévios, no Data Warehouse ou nas fontes de dados. Por exemplo: podemos definir que a Margem Bruta = Venda - Custo do produto, sem a necessidade de calculá-lo previamente.

Ou podemos definir em uma dimensão de tempo como calcular o acumulado do ano, ou comparar os resultados do mês atual com o anterior, ou ainda dois meses de anos diferentes. Esse tipo de análise é muito comum em aplicativos de negócios de Análise de Dados, o que torna muito fácil criar essas visualizações. Agora que você sabe como os dados são estruturados dentro das aplicações, tenho certeza que você terá muitas ideias para acompanhar seus negócios de novas e inovadoras perspectivas.