



Banco de dados relacional e Big Data

Proposta de resolução

Autoria do Desafio Profissional: Cassio Rodolfo Aveiro da Silva

Leitura crítica: Flavio Fiuza

Administrar um jornal de grande divulgação e circulação requer uma excelente gestão de um grande volume de dados e dos processos de negócio relacionados, como as campanhas de marketing, por exemplo. Dados sobre logística e distribuição e recursos financeiros e humanos também são pontos importantes a serem trabalhados.

Iniciando a busca pela resolução do problema da The Journal em gerir adequadamente seus dados, observa-se que os vários repositórios e bancos de dados locais surgiram pelas dificuldades de acesso impostas pelo *mainframe*. Extrair dados desse sistema e convertê-los aos formatos requeridos pelos variados sistemas internos da empresa se mostram uma tarefa árdua, além de morosa. Se forem dados de utilização recorrente, ou seja, que precisam ser acessados várias vezes e diariamente durante as operações dos processos de negócio, torna-se muito mais prático salvá-los (depois de previamente tratados) em um local de fácil acesso. Assim, os repositórios locais passam a ser uma prática comum que traz mais agilidade nas operações empresariais.

O sistema atual baseado em *mainframe* é uma tecnologia desatualizada de gestão de dados. Além disso, todo sistema, por mais eficiente que seja, precisa vir ao encontro das necessidades, características e peculiaridade da empresa que o utilizará e, acima de tudo, estar atualizado em função das demandas atuais. Se essa relação não estiver coerente, não haverá sucesso.

Deve-se utilizar um sistema de gestão de banco de dados (DBMS) capaz de monitorar o desempenho de toda a empresa no que tange ao acesso e à transformação de dados em informações pertinentes e com grande significância. A The Journal necessita reduzir custos de campanhas de marketing não direcionadas adequadamente e melhorar o atendimento ao cliente. Logo, olhando a gestão de dados nas dimensões de pessoas, organização e tecnologia, algumas reflexões devem ser feitas.

Começando com questões sobre organização, é necessário centralizar o armazenamento de dados em um único repositório e promover alterações na estrutura organizacional que façam com que todos os setores da empresa

tenham acesso a esse banco de dados central, integrando os processos de negócios. Dessa forma a The Journal poderá confrontar facilmente os dados referentes aos futuros e atuais clientes, consumidores que não querem ser importunados, aos que preferem o jornal impresso e àqueles que optam pela versão digital. Para isso, será preciso selecionar no mercado uma tecnologia de informação que permita essa integração entre sistemas e aplicações.

Quanto às pessoas, deve-se compreender que, apesar de todos os benefícios advindos com uma atualização na gestão de dados, nem todos os usuários internos da empresa migrarão imediatamente para um novo sistema. Os colaboradores já estão acostumados a extrair dados do *mainframe*, manipulá-los e salvá-los em seus próprios bancos de dados locais e tendem a continuar a trabalhar dessa mesma forma. Portanto, você deverá realizar projetos educacionais, como treinamentos e capacitações, principalmente nos setores de marketing e vendas, para que os funcionários tenham o discernimento sobre a importância da gestão eficiente de dados em toda a empresa, as facilidades que serão obtidas com a reunião de todos os dados da empresa em um só lugar, e, acima de tudo, a maior facilidade em garantir a segurança de dados e de informações sigilosas. Neste ponto, é importante que você pesquise e desenvolva uma política de informação que regule a manutenção, distribuição e uso dos dados e informações na The Journal, visando a segurança e evitando a ocorrência de inserção de dados inexatos, incompletos ou inconsistentes que possam gerar sérios problemas operacionais e financeiros. É preciso garantir segurança aliada à qualidade dos dados.

Por fim, no que tange à tecnologia, haverá duas opções de tecnologia de informação que podem substituir o ultrapassado *mainframe*, ambas de maneira bem eficaz:

1ª opção

Utilização de um banco de dados relacional (BDR) aliado a um sistema de gestão de banco de dados adequado. Com essa proposta de solução, haverá uma fonte única de dados que atenderá a múltiplas aplicações ou setores por meio do DBMS, que possibilitará, sem grande esforço, transformar os dados físicos em informações lógicas significativas aos interessados. O BDR terá o benefício de utilizar a Linguagem Estruturada de Consulta (SQL), que contribui para a integração de sistemas existentes e do próprio banco de dados ao DBMS. O banco de dados relacional permitirá a criação de tabelas (entidades) de dados principais por meio da combinação de outras entidades (relações entre dados), que reunirão os dados mais importantes e mais utilizados diariamente, evitando, sobretudo, redundância de dados. Ajudará também a compreender as relações existentes entre dados de diferentes processos de negócios por meio dos diagramas entidade-relacionamento, com o surgimento de novas oportunidades antes não visíveis. O único, porém, no uso de banco de dados relacional está no fato de ser limitado a dados estruturados e

apresentar uma capacidade reduzida frente a uma alta demanda de dados, ou seja, para um grande volume de dados, as funções de identificação de padrões e relações entre os dados deixam de promover transações rápidas em solicitações de consultas pelos usuários.

2ª opção

Utilização de Big Data aliado a um *data warehouse*. Neste caso, utilizar-se de Big Data seria muito interessante em função do grande volume de dados a serem processados de maneira rápida, uma vez que há 3,5 milhões de leitores clientes do The Journal, fora os clientes potenciais. Além disso, a prática de Big Data permite o armazenamento de informação em nuvem, eliminando investimentos em grandes computadores para a organização, manipulação e armazenamento de dados que, na maioria das vezes, encontram-se não estruturados ou, no melhor dos casos, semiestruturados. Isto já resolveria também a questão de segurança dos dados referente às falhas físicas da infraestrutura de tecnologia de informação e qualidade das informações. Para que a concepção de Big Data funcione, são necessárias ferramentas importantes de gerenciamento de dados. Aqui é que entra o *data warehouse*, o repositório local único que centralizaria todos os dados da empresa provenientes das várias fontes presentes em Big Data. Facilitaria o acesso aos dados e traria, de modo geral, poucas mudanças comportamentais nas operações dos funcionários que já estão acostumados a acessar um banco de dados local. O *data warehouse* potencializa a aquisição de dados externos à empresa (já possibilitado por Big Data) e que são de uso recorrente, como cotações monetárias e taxas cambiais que sempre devem ser apresentadas no jornal de forma atualizada. A principal vantagem do *data warehouse* é que ele pode extrair dados de múltiplos sistemas em operação dentro da empresa, combinando-os com os originados por fontes externas e permitindo gerar relatórios de gestão e análises. Estes são fundamentais para as tomadas de decisões assertivas e ficam também armazenados no próprio *data warehouse*. O único ponto crítico nessa opção de solução é que, se não planejada adequadamente, pode gerar altos custos com implementação e manutenção.

Conclusão

Com as decisões direcionadas sobre as questões de segurança e qualidade de dados, capacitação e estímulo aos colaboradores à atualização da estrutura organizacional referente à gestão de dados de clientes e futuros consumidores, o ponto principal está em escolher qual será a tecnologia a ser utilizada: baseada em práticas relacionais ou em Big Data? Frente à complexidade exigida na análise dos dados, percebe-se uma não estruturação, ou, no caso, uma semiestruturação, pois deve haver um cruzamento de dados a fim de manter os clientes atuais sem incomodá-los, fornecendo-lhes os serviços preferidos e ao mesmo tempo atrair mais leitores, analisando fontes e tendências externas no

país e no mundo. Assim sendo, o BDR poderia ser utilizado, mas perderia sua agilidade relacional frente ao imenso volume de dados, o que era seu ponto forte. Portanto, deve-se optar pela segunda opção, que é a utilização de Big Data e *data warehouse*, que permitirá um acesso e processamento de dados muito maior. Além disso, as relações entre dados, que poderiam ser realizadas pelo banco de dados relacional, são desenvolvidas pelos *data warehouse* e apresentadas de maneira clara e objetiva por meio de relatórios, facilitando muito as decisões estratégicas a serem tomadas para melhorar a eficiência operacional do jornal. Apesar de um possível custo de investimento, a escolha por Big Data e *data warehouse* produzirão economias advindas de ganhos de eficiência, simplificação de processos e redução de queixas dos consumidores, o que justificaria seu uso. É uma opção que pode também ser aplicada em conjunto com variados tipos de DBMS.

