Big Data vs Data Warehouse

Diferentemente de um Data Warehouse, o Big Data vai além da consolidação de informações, pois é utilizado principalmente para o armazenamento e processamento de qualquer tipo e volume de dados com um volume que potencialmente cresce exponencialmente. De certo modo, *Data Warehouse* (DW) e *Big Data* (BD) possuem o mesmo propósito, isto é, prover suporte a tomada de decisões, explorar dados para a identificação de padrões, geração de estatísticas e indicadores de desempenho (SALINAS e LEMUS, 2017).

As diferenças entre DW e BD estão na natureza dos dados, nos usuários a que são destinados e nos procedimentos e ferramentas para aquisição, armazenamento e análise de dados. *Big Data* foca principalmente na exploração de dados brutos, como dados não estruturados e não repetíveis, não susceptíveis a agregações e tratamentos sistemáticos. Por isso, geralmente, usuários de *Big Data* são mais especializados (por exemplo, cientistas de dados), que com a utilização de ferramentas, técnicas e algoritmos especiais conseguem identificar padrões que resultam em conclusões valiosas. *Data Warehouses* são focados em dados estruturados e alguns tipos de dados não estruturados, que precisam ser pré-processados antes de serem disponibilizados aos usuários finais (que podem não ter o conhecimento necessário de mineração de dados ou outras conhecimentos específicos), deste modo, estes podem fazer análises independentemente da origem dos dados, tipo de armazenamento, arquitetura, ferramentas e algoritmos específicos (SALINAS e LEMUS, 2017).

Para Salinas e Lemus (2017), a integração de conjuntos de dados heterogêneos pode ser a principal diferença entre um *Data Warehouse* e uma aplicação de *Big Data*. Em um DW, o propósito de uma integração é a obtenção de uma visão uniforme da organização, enquanto que em uma aplicação de BD, a integração não consiste em seu objetivo final. Neste último caso, conjuntos de dados não estruturados não passíveis de integração devem ser mantidos em seu formato original, permitindo a possibilidade de usos futuros, não previsíveis para o momento.

Em seu artigo, Salinas e Lemus (2017) concluem que aplicações de Big Data não constituem uma evolução para *Data Warehouses*. Na realidade DW e BD são complementares e podem ser integrados para o compartilhamento não apenas de dados, mas também de armazenamento e demais recursos computacionais.

Salinas e Lemus (2017) alertam que a demanda por soluções rápidas e a versatilidade de algumas ferramentas têm levado ao desenvolvimento de projetos de DW sem a utilização de metodologias e frameworks adequados. Após o investimento de tempo e recursos financeiros, organizações têm se visto presas em “soluções” de DW/BI que não cumprem as expectativas iniciais, inflexíveis a mudanças, difíceis de manter e escalar. Em relação a soluções de Big Data, existe atualmente todo um ecossistema de tecnologias, não necessariamente integradas em uma única plataforma, que pode aumentar a complexidade de desenvolvimento de novos projetos.

Big Data substitui a necessidade de um Data Warehouse?

Como já mencionado por Salinas e Lemus (2017), Big Data e Data Warehouse são complementares. Esta é também uma visão da Oracle (2022), ao afirmar que muitos veem *big data* como uma extensão integral de seus recursos existentes de *business intelligence*, plataforma de *data warehousing* e arquitetura de informações.

Para Anand (2019), uma solução de *Big Data* é uma tecnologia, enquanto que *Data Warehousing* seria um conceito arquitetônico em computação de dados. Uma organização pode ter diferentes combinações, com apenas soluções de *Big Data* ou de *Data Warehouse,* ou ambas simultaneamente, a depender dos seguintes fatores: Estrutura dos dados na origem; volume de dados; velocidade com que os dados deverão estar disponíveis para análise; e nível de conhecimento dos usuários dos dados.

Do ponto de vista do usuário final, em um DW, os dados armazenados estão sempre prontos para serem consumidos, mas são limitados aos dados já armazenados. Os dados em um DW estarão orientados por assunto/processo de negócio, portanto mais facilmente entendíveis por usuários finais. No entanto, se os usuários precisam de dados não presentes em um DW (como por exemplo, dados de redes sociais ou de algum log), a inclusão poderá não ser ágil devido ao processo necessário para coleta, tratamento e ingestão de dados de forma estruturada e organizada. Tais dados poderiam ser direcionados a uma solução de *Big Data*, permitindo com que os usuários pudessem acessá-los em formato nativo, mas que precisariam de transformações para a elaboração de relatórios. A recuperação de dados presentes em soluções de *Big Data* é difícil, pois os dados armazenados são desestruturados e desorganizados (ANAND, 2019).

Revisão #18
Criado 2022-03-08 12:48:15 UTC por FLAVIO LOPES DE MORAIS
Atualizado: 2022-12-14 09:02:06 UTC por FLAVIO LOPES DE MORAIS