

# Data Warehouse

*Data Warehouse* (DW) consiste em um sistema para armazenamento de dados originados de múltiplas fontes, especialmente estruturados para consulta e análise. Um DW busca a criação de uma fonte de dados padronizada, confiável e de acesso simplificado, para apoio à tomada de decisão (Kimball e Ross, 2013).

De acordo com o DAMA (2017), um DW é a combinação de dois componentes principais: (a) um banco de dados integrado para apoio à tomada de decisão e (b) o conjunto de softwares relacionados, utilizados para coletar, limpar, transformar e armazenar os dados a partir de várias fontes de origem.

De acordo com Anand (2019), *Data Warehouse* pode ser considerado um modelo arquitetônico para armazenamento de dados estruturados. Não sendo portanto uma tecnologia em especial.

Segundo Inmon (2005), um DW consiste em uma coleção de dados orientados por assunto, integrados, não voláteis e variantes no tempo, com o intuito de prover suporte à tomada de decisão. Na arquitetura de DW proposta por Inmon (2005), a não volatilidade de um DW deve ser garantida por meio do que é chamado pelo autor de *snapshots*. De acordo com Inmon (2005), uma vez inserido em um DW, o dado não poderá mais ser atualizado. No entanto, em seu próprio livro, são demonstradas algumas opções de alteração de dados históricos, como por exemplo, a correção de um valor incorreto de saldo bancário histórico de cliente. Este aspecto de não volatilidade absoluta conflita com a proposta de arquitetura de *Data Warehouse* de Kimball e Ross (2013). Na arquitetura de Kimball e Ross (2013), a manutenção da história em um DW é feita por meio da técnica *Slowly Changing Dimensions* (SCDs). Esta técnica assume a possibilidade de modificação de dados históricos em um modelo dimensional, possibilitando este controle a nível de atributos.

De acordo com Khine e Wang (2018), um *Data Warehouse* ser não volátil significa que os dados permanecem inalterados entre as cargas de dados. Diferenciando-se dos dados transacionais de sistemas que podem ser alterados a todo instante.

Existem diversas abordagens para a construção de *Data Warehouses*. Kimball e Ross (2013) propõem uma abordagem de construção de um DW que integre toda a organização (*Enterprise Data Warehouse - EDW*). De acordo com os autores, a construção deste tipo de DW é fundamental para a governança de dados. Ainda de acordo com Kimball e Ross (2013), a ausência de um DW institucional como plataforma de BI, alinhado à uma boa governança de dados, leva à perpetuação de silos de dados similares entre departamentos, mas com versões da verdade ligeiramente diferentes.

A UFLA utiliza a arquitetura de *Enterprise Data Warehouse* proposta por Kimball e Ross (2013). Esta escolha deve-se ao fato desta ser uma arquitetura já consolidada, amplamente aceita e utilizada no mercado.

O EDW é construído com a técnica denominada Modelagem Dimensional, que, segundo Kimball e Ross (2013), trata-se de uma abordagem amplamente aceita para consolidação de dados

analíticos por abordar dois requisitos de forma simultânea:

- Apresenta dados em formatos entendíveis por usuários de negócio;
- Organiza os dados de modo a otimizar o desempenho de consulta.

Kimball e Ross (2013) mencionam ainda as seguintes técnicas existentes para a construção de *Data Warehouses*:

- *Independent Data Mart Architecture*;
- *Hub-and-Spoke Corporate Information Factory Inmon Architecture*;
- *Hybrid Hub-and-Spoke and Kimball Architecture*.

Kimball e Ross (2013) demonstram que as abordagens mencionadas acima possuem grandes desvantagens em comparação à abordagem de EDW com Modelagem Dimensional. Para mais detalhes sobre essas desvantagens consulte as páginas 26, 27, 28, 29 e 30 (KIMBALL e ROSS, 2013).

Um EDW considera os seguintes princípios para a sua construção:

- Utiliza Modelagem Dimensional do tipo *Star Schema*;
- Armazena dados na menor granularidade possível (embora possa também armazenar agrupamentos e sumarizações, realizadas a partir dos dados granulares);
- Os fatos são orientados a processos de negócio, e não a setores ou departamentos específicos;
- Utiliza Dimensões Coformadas.

A Figura abaixo apresenta os elementos chave para a arquitetura Kimball de DW/BI (KIMBALL e ROSS, 2013):

De acordo com Khine e Backlund (2018), como um *Data Warehouse* possui uma estrutura fixa, com processos de extração, transformação e carga de dados muito bem definidos, este possui um aspecto forte de gestão de governança de dados.

