

Linhagem de Dados

De acordo com Kimball e Ross (2013), a linhagem de dados descreve as origens e as etapas de processamento que produziram cada elemento de dado em uma tabela. Esta linhagem deve ser armazenada no *Data Warehouse / Data Lake* por meio dos metadados associados a cada elemento. Essa linhagem é explicitamente exigida por certos requisitos de conformidade, mas deve fazer parte de todas as situações de arquivamento.

Segundo Barbieri (2019), Linhagem de Dados significa o entendimento de cada passo ao longo dos processos, observando quais dados e metadados entraram em cada bloco de processamento, quais dados saíram, quais foram os processamentos efetuados, regras aplicadas, etc. Deste modo, quando houver um problema específico no fluxo de processamento de dados, a linhagem poderá auxiliar na depuração e identificação do problema. Conhecer os detalhes deste caminho é fundamental para a depuração de erros que possam surgir em um ponto mais adiante do fluxo, mas que podem ter sido produzidos em estágios anteriores.

Figura 1: Exemplo de visualização de um grafo de exibição de linhagem de dados.

