



***Big Data Real-Time Analytics Com Python e Spark 3.0***

# **Big Data Real-Time Analytics Com Python e Spark Versão 3.0**

## **Estudo de Caso 4 Definição do Problema e Fonte de Dados**



Uma empresa internacional de comércio eletrônico (E-commerce) que vende produtos eletrônicos deseja descobrir informações importantes de seu banco de dados de clientes.

Os produtos ficam armazenados em um armazém na sede da empresa. Após concluir a compra no web site da empresa, o cliente recebe o produto em casa, em qualquer parte do mundo. Os produtos são enviados de Navio, Avião ou Caminhão, dependendo da região de entrega.

Em cada compra o cliente pode receber um desconto dependendo do peso do produto comprado. Cada cliente pode fazer chamadas ao suporte da empresa no caso de dúvidas ou problemas e após receber o produto o cliente pode deixar uma avaliação sobre a experiência de compra. O único dado pessoal sobre o cliente que está disponível é o gênero.

Nosso trabalho neste Estudo de Caso é criar novas variáveis (novos recursos) que nos ajudem a compreender os dados de forma mais simplificada e por diferentes perspectivas.

Trabalharemos com dados fictícios que representam dados reais de uma empresa de E-Commerce. Os dados estão disponíveis ao final do capítulo.

O resultado deste trabalho será usado no próximo capítulo.