Business Use Case: E-Commerce Sales Analytics with Hive Partitioning and Bucketing

# Problem Statement

An e-commerce company processes massive amounts of sales data from customers across different regions and countries. The data is queried frequently for analytics purposes, such as:  
1. Daily Sales Reports by country and region.  
2. Product Sales Trends to identify the top-selling products in specific regions.  
3. Customer Purchase Patterns for customer segmentation and targeted marketing.  
  
The dataset is large (millions of records), and querying it without optimization leads to slow performance. The company wants to:  
- Optimize queries for specific date ranges (daily sales reports).  
- Speed up queries for product categories.  
- Use partitioning and bucketing to manage this large dataset efficiently in Hive.

# Sample Data

## Sales Table (sales\_data)

Columns:  
- sale\_id: INT  
- product\_id: INT  
- product\_category: STRING  
- customer\_id: INT  
- sale\_amount: FLOAT  
- sale\_date: DATE  
- country: STRING  
- region: STRING  
  
Sample Data:  
1, 101, Electronics, 2001, 500.00, 2023-08-01, US, North America  
2, 102, Electronics, 2002, 300.00, 2023-08-01, US, North America  
3, 103, Furniture, 2003, 700.00, 2023-08-02, UK, Europe  
4, 104, Furniture, 2004, 800.00, 2023-08-03, UK, Europe  
5, 105, Clothing, 2005, 200.00, 2023-08-03, IN, Asia  
6, 106, Clothing, 2006, 600.00, 2023-08-03, IN, Asia