

# Module 04: Spark ML Applications

## I. Logistic Regression Using Spark

- Case Study: Online Advertising
  - Click-through rate is a very important metric for evaluating ad performance.
  - **CTR = Clicks/Impressions**
  - CTR is basically rate how many users clicked on Ad with respect to how many times the Ad was displayed
  - As a result, click prediction systems are essential and widely used for sponsored search and real-time bidding.

## II. K-Means Clustering Using PySpark

- Case Study: Music Artists Clustering
  - Problem Statement:
    - Find Cluster or Similar music artist from the data based on thier popularity in terms of how many times people have listened to their song.
    - Can be used for recommendation: similar artist song to users
    - Useful for monetization and business point of view: Exclusive launch of songs on the platform [profit sharing]
    - Solve cold start problem: Categorizing new artist songs in a cluster based on the features