# Assignment II

# Apache Kafka – Spark Integration

# Note:

# *This is a take-home assignment to be carried out by the group of learners.*

# *There are three programming exercises - requiring only one dataset to be used – on Apache Kafka – Spark integration.*

# *The learner may use any programming language as the programming interface, while Java or Python is recommended.*

# *You may consult / discuss with other learners peripheral aspects such as the environment but not on solving the specific problems in terms of design or implementation.*

# End of Note.

# Learning Outcomes:

# Ability to write programs that performs data processing using the Apache Kafka – Spark frameworks

# Secondary outcome is getting familiarity with custom implementation of data pre-processing, join and other operations on the data.

# Assume that you are working as analyst for “MyMall”, a supermarket chain. The mall has collected some interesting characteristics of customers who had visited the mall earlier. (Refer the attached Mall\_Customers.csv file for the same). The marketing management is planning a campaign to increase the sales of a new product. As a part of this exercise, before opening the campaign for masses, the team want to apply a marketing offer on the customers who are roaming around a mall (vicinity of 100 meter) or who are within the mall. You are supposed to carry out following tasks programmatically to help to roll out the coupons to the identified customers.

# Exercise 1:

# Assume the customers are roaming around mall or are present within the mall and the sensor / application placed on their mobile device is helping to capture these movements. Write a custom code that will mimic the movement of the customers and make it available in appropriate format to a Kafka Topic designated for the same purpose. Make necessary assumptions about the format of movements of customer.

# Exercise 2:

# Now you have to wear your hat of analyst! Attached customer dataset provides some additional information about the customers like age, income etc. You are supposed to correlate the customer stream data with this dataset and propose the discount offers to the customers who are near to the mall. Make suitable assumptions while designing the offers. Briefly explain the strategy used to design the offers.

# Exercise 3:

# Now it’s time to roll out these offers to the identified customers. For that purpose, you need to process the customer movements received through the Kafka topic. You can think of following activities to be carried out on the stream of customer data within the Apache Spark.

# Preprocessing on the incoming messages

# Determining the distance of customer from the mall

# Filtering out the customers based on certain distance criteria

# Join the customer stream with the customer dataset

# Propose a suitable discount offer/coupon based on the enriched data stream

# Provide this offer details to a designated Kafka topic so that downstream application can make use of it to actually send it through SMS.

# References

# Kafka / Spark Lab sheets made available on the course website

# Dataset

# 

# ======================END==================