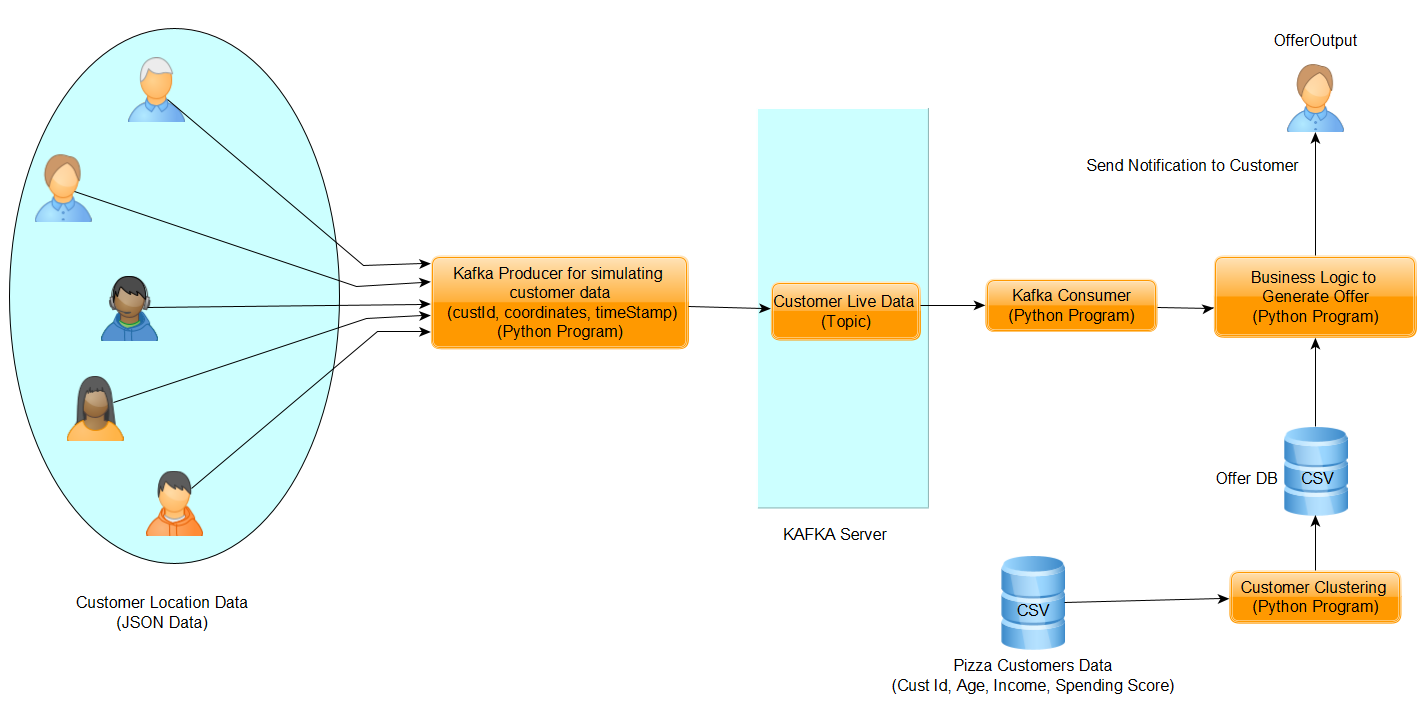
1. **Streaming data pipeline architecture**
   * Firstly, we will read data from Pizza Customer Data csv
   * The csv will help us to understand the customer data which contains its income and its spending score
   * With above mentioned data we used K-means clustering (without library) to form clusters based on 5 segments
     + Low-income high transaction
     + Low-income Low transaction
     + Medium income medium transaction
     + High income low Transaction
     + High income high transaction
   * Based on above 2 data points i.e., Income and spending score we decided to create 5 clusters and stored it in offer DB
   * Offer DB is the output result of our clustering which will help to make business decisions.
   * When customer enters the mall, producer will continuously read and transfer the customer id, their geographical location and timestamp data to Kafka topic. Here we have stored sample coordinates in Customer Location data.
   * This data will be saved at Kafka server topic which will be read by consumer.
   * Our Kafka consumer will constantly keep receiving the data from Kafka server and it will send the required geographical information and customer id to our heart of the system i.e., business logic
   * Our business logic checks if the customer exists in our data base and based on his income and spending score reads the offer to be assigned to the customer.
   * As the user is in proximity (less than 1 km distance) to our restaurant our business logic sends notification to the customer based on the discount generated from offer DB. We are saving the data in OfferOutput .csv
2. **Data Flow**



1. **Business Logic Used**

* This program reads data from Kafka topic such as customer Id, latitude, longitude, timestamp.
* Calculate distance of customer from a mall.
* If the distance between customer and mall is less than 1 km, fetch the customer data from Pizza\_customers csv file.
* Fetch the data from Offer csv file to determine the cluster, from which customer belongs to, based on income and spending score
* Generate offer if customer belongs to one of the clusters, calculate offer for the customer based on cluster nomenclature
  + **Bronze** :15% discount rate enabled them to be promoted to silver tier
  + **Silver**: We encourage them to give 5% discount as they are already under high transaction, and higher discount won’t affect the business much as their income is less
  + **Gold**: 15% discount rate can be considered as an average rate to promote them to order more from our restaurant
  + **Diamond**: We plan to give them our premium 25% discount rate
  + **Platinum**: We plan to give them least discount rate of 5%
* The generated output data saved in OfferOutput csv file

1. **Component Used**
   1. Kafka : Kafka is primarily used to build real-time streaming data pipelines and applications that adapt to the data streams. It combines messaging, storage, and stream processing to allow storage and analysis of both historical and real-time data.
   2. Zookeeper: ZooKeeper is used in distributed systems for service synchronization and as a naming registry. When working with Apache Kafka, ZooKeeper is primarily used to track the status of nodes in the Kafka cluster and maintain a list of Kafka topics and messages.
   3. Database : CSV file used to save data
2. **Clustering Nomenclature**

On basis of available data, we have divided our customer base in 5 categories

1. **Bronze**

* This category represents customer having low income and low transaction frequency record.
* We plan to give them our premium 25% discount
* 25% discount rate will enable them to be promoted to silver tier
* Giving them 25% discount will make more chances of them frequently order from our restaurant since their spending capacity is low because of low income
* Giving attractive discount of 25% will encourage them to spend and get upgraded to silver tier
* Our main target would be to convert all bronze customers to silver tier

1. **Silver**

* This category represents customers who have low income but have high transaction frequency record
* We encourage them to give 5% discount as they are already under high transaction, and higher discount won’t affect the business much as their income is less
* So, giving those customers higher discount won’t benefit the business as they are already our frequent customers
* For Low-income customers silver tier is ideal tier since their spending capacity is quite low

1. **Gold**

* This category represents customers who have medium level income and transaction frequency
* We plan to give them 15% discount
* Our focus for this tier is to increase their transaction frequency to a larger extent
* For gold tier users, their spending capacity can be considered as average
* Thus 15% discount rate can be considered as an average rate to promote them to order more from our restaurant
* This tier cannot be promoted to other tier as their income capacity comes under average range

1. **Diamond**

* This category represents customers having high income but low transaction frequency
* We plan to give them our premium 25% discount rate
* Consumers in this tier have the capacity to spend, but by giving premium discount price will attract them to purchase more from our restaurant
* Our target is to make them familiar and used to our restaurant thus promoting Diamond tier to Platinum tier
* Diamond tier is the most important tier of all in terms of increasing revenue as they have the spending capacity
* Thus, by giving them our premium 25% discount rate will leverage them to purchase more thus increasing our transaction frequency

1. **Platinum**

* This category of customers represents high level income with high transaction frequency.
* We plan to give them least discount rate of 5%
* Since these customers already have good transaction frequency with good spending capacity, we don’t need to give them more discount
* This tier is considered to be most revenue generated tier
* To preserve customers of this tier, we need to make sure we have consistent service and quality of our restaurant

1. **Submission Files:**
   * 1. CustLocationData : Jason File to store customer location data.
     2. Cust\_Clustering : Python program to cluster data using K Means without library
     3. Offer DB : This is CSV, it contains the clustered data. This is output of cluster program
     4. pizza\_customers : CSV i.e. given data
     5. ProducerToSimulate : Kafka Producer Python program which takes customer location data and send to Kafka Server. Also it saves the output in SimulatorOutput txt file.
     6. Simulator Output: Text file which containsOutput of Kafka producer
     7. SPA Assignment 2 Group192 : Doc file which Contains data flow and other documentation.
     8. ConsumerReadProcess : Kafka Consumer python program, this also contains business logic
     9. OfferOutput: CSV file. This is the output, which shows which customer got the offer.