# Review Paper: Big Data Analytics based Mobile Advertising

Ms Kawaljit Kaur, Dr Sandeep Sharma
*CET Department*
*GNDU Amritsar, Punjab, India 143001*

***Abstract*** **-** Mobile networks and Internet of things are generating endless amount of data these days. Large and complex data are generated but it is very difficult to deal with that much amount of data with traditional methods of data processing. The rapid increase in mobile computing and wireless networks have created new opportunities um mobile advertising. Big data analytics is effective approach to extract the useful information and knowledge from that much amount of data coming from mobile devices. In the paper various big data technologies and tools used in various ad recommendation frameworks in big data are analysed and reviewed.

## I. INTRODUCTION

Large amount of data is generated and arriving in such a complex and fast way need to be managed store and analysed. Sensors, log data of machines, data storages, public web, social media, business apps, media, archives, and numerous other types of technologies are creating and capturing data continuously in large quantities. This data is very useful in targeting the customers for ad recommendation on mobile phones. There are various reasons for increase in digital marketing like our time spent online is increasing, decline of publishers' offline business, invention in digital marketing technology and creative.

The use of mobile devices is increasing unprecedentedly as indicated in [1],"every day quite one million new mobile devices are activated worldwide. Similarly, Apple in Sept 2014, proclaimed that "it had sold over ten million iPhone 6's within the initial 3 days of it being offered. This is often just one million quite the over nine million iPhone 5c's and 5s' that it sold in 2013. But the complete potential of the internet is yet to be released. The different mobile users exhibit different demographics, personal preferences, behaviour, social presence and location usage. In order to engage with customers in a meaningful way, one must analyse the large and numerous datasets. The social and commerce activities that consumers interact on their mobile phones are redefining mobile advertising. Mobile advertising based on big data analytics face many challenges like unstructured and structured mobile data, the exclusivity of the data, different mining techniques for mobile user data[1]. With big data, businesses and organizations can learn a lot of concerning their business and operations and that they will translate that knowledge into improved decision making processes and improved performance. Big data is different from regular data having three characteristics, generally called 3 v's: Volume,, Velocity and Variety.

**A. Volume:** It is quite common to own Terabytes and Petabytes of the storage system for enterprises. As indicated in [2], it was estimated that around one billion gigabytes of data are created from different sources day by day in 2012. This number is doubling every 40 months. Sources of data are Internet, mobile devices, smartphones and sensors.

**B. Velocity:** Organizing, accessing and processing the huge amounts of latest data created every second as it is collected to be included in the decision making in real-time applications is usually considered a complex and difficult technical challenge. Velocity is one amongst characteristics of Big Data that talks concerning the high data rate at which it's being generated. Various applications based on data rate are Batch, Real Time, Interactive, Streaming[3].

**C. Variety:** Data is usually collected from different sources like messages, images and videos posted to social networks, readings from sensors, business transactions, and economic and political news.

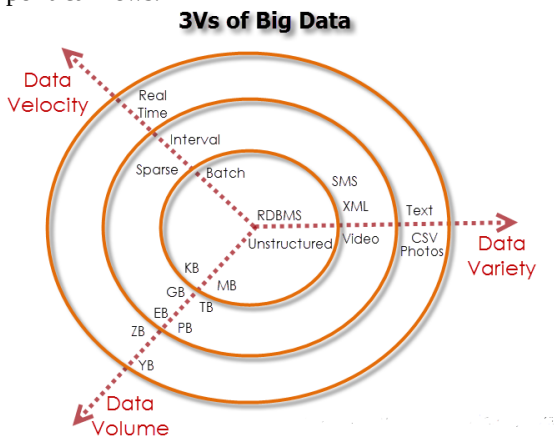


Figure 1: 3 v's of big data

For targeted advertising and marketing+ many companies have begun to use big data technology. However some ad placement companies are building their own system for better ad targeting.

## II. TOOLS AND TECHNOLOGY USED FOR BIG DATA ANALYTICS

There are number of tools used for big data analytics. We have discussed some of them as follows:

**A. Hadoop** - The Apache Hadoop package library permits for the distribution process of enormous information sets across clusters of computers whereas victimization easy programming models. It will rescale from single server to several machines, each of them having their own storage and process. Here there's no ought to depend upon hardware to attain high-availability. The library itself is meant to observe and handle failures at the application layer; it delivers a highly-available service on top of a cluster of computers, every of which can be at risk of failures. Hadoop is a distributed processing framework supported Java effective in data-intensive analytics [4] . Many methods for data analysis can be implemented utilizing Hadoop for processing, analytics, and storage. Hadoop MapReduce is restricted to instruction execution of 1 job at a time [4]. In addition, it offers parallel computing that contributes to high performance and efficiency for big data analytics projects. Hadoop provides its own file system called HDFS (Hadoop Distributed File System).

File systems managing the storage across machines in a network are referred to as distributed file systems. When we deploy our data on HDFS, Hadoop divides all the data in different clusters and performs parallel opreations. It also keeps number of copies of data in case of hardware failure. A Hadoop distributed file system cluster works in a master-slave pattern: **a Name-Node (master)** and **a number of Data-Nodes (slaves)** [5].

A user can buy an entire bunch of commodity hardware to run Hadoop on single machine. Running Hadoop on a totally configured cluster leads to execution of a group of daemons or resident programs, on the various servers in network [5]. These daemons include:

**NameNode:** Namenode is the master node which provide the meta data storage of the shared file system and leads the slave DataNode daemons to perform I/O tasks.

**DataNode:** Every slave machine in cluster hosts a DataNode daemon to do reading and writing HDFS blocks to real files on the local file system.

**Secondary NameNode:** At each interval secondary namenode gets snapshots of the HDFS metadata from NameNode.

**JobTracker:** The main task of JobTracker is to determines which files to process, assigns nodes to tasks, and keeps the track of all tasks.

**TaskTracker:** TaskTrackers manage tasks executing on each slave node and executes the individual tasks assigned by the JobTracker.

The working of Map Reduce function with all the components is depicted in diagram given below:

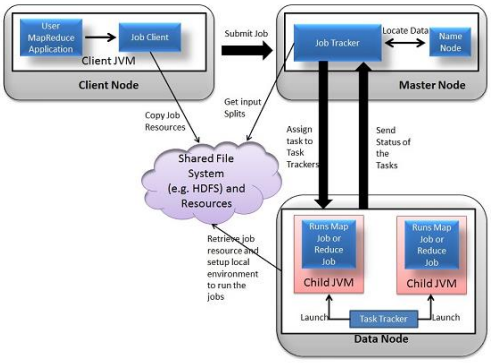


Figure 2: Working of Map Reduce

**Map Reduce Function:** The Map-Reduce is proposed by Google. It a parallel data processing model which divides the computation in two functions, Map function and Reduce function. The steps to process Map reduce programming models would be as follows:

| | |
|---|---|
| 1. | Fetch the input data |
| 2. | Split the data into partitioned blocks |
| 3. | Assign the blocks to Map tasks. |
| 4. | Output of the Map tasks is sorted |
| 5. | The sorted data is reduced by using the Reduce function. |
| 6. | Store the output in the file system |

The creators of Facebook Messaging, utilize Hadoop for their application because of its elasticity, fault isolation, low latency, and consistency semantics.

**B. Spark** - As indicated in [7], "Spark process data on memory. Spark runs 100 times faster than Hadoop". Spark is an open source project developed by Apache in order to handle big data. It provides a distributed computing platform which permits fast, efficient, fault-tolerant and scalable processing large, complex and massive data. The Spark ecosystem has some basic components such as Spark SQL, Spark Streaming, MLib and GraphX that makes it more powerful.

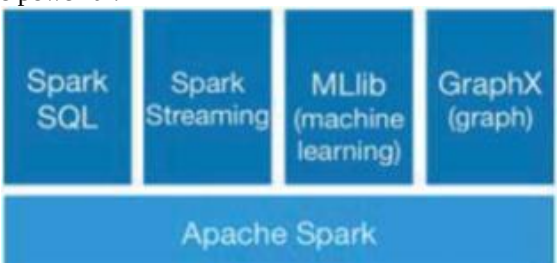


Figure 3: Spark Ecosystem [6]

Spark SQL is used to query data with SQL and provides support for structured and semi structured data. Spark Streaming facilitates stream processing, MLib is a library contains machine learning algorithms and GraphX is basically used for graph analytics. Spark is designed to use the basis of Hadoop MapReduce with some modifications that enables it to do more types of computations, which includes interactive queries and stream processing more efficiently than Hadoop MapReduce. Spark framework has its own streaming API and autonomous processes for

continuous batch processing across different short time intervals. We can consider Spark as an extension of Hadoop. Spark runs up to 100 times faster than Hadoop in certain problems; however it still uses Hadoop distributed file system. That is why most of the Big Data based projects install Spark on Hadoop to use data stored in Hadoop distributed by advanced big data applicatiobs. Resilient Distributed Datasets (RDDs) are distributed memory abstraction facilitates programmers to perform in- memory computations on big clusters to take the advantage of the speed [6].

**C. Storm** - As indicated in [7], According to Eclipse Public License 1.0, Storm is a distributed and fault-tolerance real-time computing system. Strom can easily compile and expand complicated real-time computation in a computer cluster. Storm can ensure that each message will be quickly disposed. Furthermore, any program language can be used for development. Storm has following main features:

1. Storm is a simple programming model
2. A service framework support hot deployment.
3. Storm is highly fault-tolerance.
4. Computations in storm are done in parallel way.
5. It is quite Reliable in message handling.
6. "Local Mode" simulation

The Storm Cluster consists of one main node and many working nodes. A daemon process in Storm called "Nimbus". Nimbus runs on main node. Its main task i to allocate codes, arrange tasks and detect errors. Each working node has a daemon process called "Supervisor". Supervisor's main task to monitor, start and stop working process. Zookeeper handles the coordination work between Nimbus and Supervisor as shown in Figure given below:

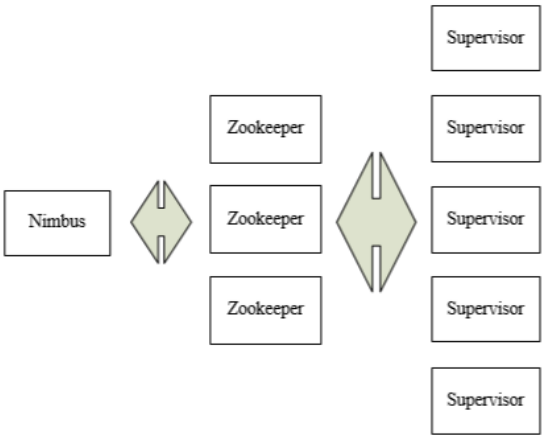


Figure 4: Structure of Storm Cluster

Zookeeper is the subproject of Hadoop, and it aims at coordinate works in large-scale distribution system. The Storm Cluster is similar with Hadoop, where Job Tracker is our Nimbus, and Task Tracker is our Supervisors. Storm is an event processing system that has the flexibility to discover necessary event occurrences It is the processing system that Jones utilizes to discover crucial events through the processing of feeds of Twitter [4].

**D. Flume** - Flume is a data ingestion tool having a very important role in real-time analytics [4]. Flume is one of the tools that prepare a distributed, reliable, and available service for efficiently importing data coming from different sources such as Twitter, Facebook etc. It is a highly reliable, distributed, and configurable tool. Collecting, aggregating, and bringing in huge amount of data with its versatile architecture supported streaming data flows, makes it potential for Big Data frameworks to ingest data in a way that makes it easy for processing tools to reach data. Working of flume is depicted in figure:

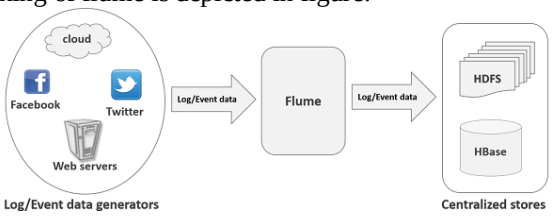


Figure 5: Flume as data ingestion tool

Flume acts similar to that of storm when applied to real time data analytics.

**E. Kafka:** Kafka is the another powerful tool for data ingestion. Kafka is a distributed streaming platform and a message broker project which provides a unified, high-throughput data feeds [4]. Characteristic of Kafka, are as follows, it is low- latency platform for handling real-time streaming data. It is, in its essence, a scalable message queue which has designed to distribute transaction log making it highly valuable for infrastructures to process streaming data.

Another big data analytics tool is S4. It is a popular framework for in-memory stream processing. It is able to handle data continuously as it arrives without terminating [4].

## III. BIG DATA ANALYTICS TECHNIQUES

There are number of Big Data techniques including cluster analysis, association rule learning, data mining, text analytics crowdsourcing, machine learning, classification, data fusion, network analysis, optimization, predictive modelling, regression, special analysis, time series analysis and others. Which technique is used depends upon various things such as the type of data being analysed, the research questions one is trying to solve and technology available. We have discussed some of them as follows:

**Predictive analytics:** It comprises of several techniques for identifies meaningful patterns of big data to predict future outcomes and asses the attractiveness of various options based on the past or historical and current data. The various predictive modelling techniques used are NNs, linear and logistic regression, SVMs, decision trees, clustering, association rules, and scorecards[8]. We can use predictive analysis techniques in all the disciplines from the prediction

of failure of jet engines based on data streams coming from thousands of sensors, to predicting customer's next moves based on what they are buying, when they are buying and what they are posting on social media sites. Predictive analysis captures the relationships in the data and uncovers patternsand. It is categorized in two ways: regression techniques (e.g.polynomial logic models) and machine learning techniques (e.g. clustering). Big data is massive and has features like heterogeneity, noise accumulation, supurious correlation ans incidental endogenity. So there is a need to develop new statistical methods for handelling big data and gain valuable insight. Some of the algorithms are discussed below:

**A. Decision Tree:** There is no need of domain knowledge in decision tree algorithm that is why it is a widely used classification technique in big data [4]. This algorithm is used to classify the attributes and decide the outcome of the class attribute. Both class attribute and item attributes are required to construct decision tree. Decision tree is a tree like structure where the intermediate nodes represent attributes of the data, leaf nodes represents the outcome of the data and the branches hold the attribute value.
The main step in the decision tree algorithm is to find the root node for the given set of data. Number of methods are there to decide the root node of the decision tree. Information gain and Gini impurity are the two main methods used to find the root node. In decision tree to decide which the data will fall is depends upon the root node. Decision trees are constructed using the training data and tested with the test data like other classification methods.

**B. Clustering:** Clustering is the technique of identifying and classifying the items in to similar groups. K-means clustering is classifying the items into k clusters based on basis of their similarity. K is denoted as the number of clusters that we have to decide before starting the clustering process. The whole process depends on the K value. That's why it is very important to choose a correct K value. The data point is grouped in to a cluster based on the Euclidean distance between the point and the centroid of the cluster. Initial clustering can be done in one of three ways.

**(i) Dynamically Chosen:** This method, is basically to choose the first K items and assign to K clusters.
**(ii) Randomly Chosen:** This method randomly select the values and assign them to K clusters.
**(iii) Choosing from Upper and Lower Boundaries:** This method will choose the values that are very far from each other and use them as initial values for each cluster.

**Steps of clustering algorithm**

Step 1: Choose the initial values using one of the above three methods
Step 2: For each additional value.
Step 3: Calculate the Euclidean distance between this point and centroid of the clusters.
Step 4: Move the value to the nearest cluster.
Step 5: Calculate the new centroid for the cluster.
Step 6: Repeat steps 3 to 5
Step 7: Calculate centroid of the cluster.
Step 8: For each value
Step 9: Calculate the Euclidean distance between this value and the centroid of all the clusters.
Step 10: Move the value to the nearest cluster.

This technique produces tight clusters than other clustering techniques. But the main disadvantage is to decide the number of clusters before starting the process [4].

**C. K-Nearest Neighbour:** Another classification technique is k-Nearest neighbour. The k-nearest-neighbour method was first defined in the early 1950s. The method is effortful when given large training sets, and did not gain popularity until the 1960s when increased computing power became available. It has since been widely used in the area of pattern recognition. Nearest-neighbor classifiers supported learning by analogy, that is, by comparing a given test tuple with training tuples that are similar to it. The training tuples are described by n attributes. Each tuple represents a point in an n- dimensional space. In this way, all of the training tuples are stored in an n-dimensional pattern space. When given an unknown tuple, a k-nearest-neighbor classifier searches the pattern space for the k training tuples that are closest to the unknown tuple. These k training tuples are the k "nearest neighbors" of the unknown tuple. When the 'k' closest points are obtained, the unknown sample is then assigned to the most common class among those k-points. In case of k=1, the unknown sample is assigned to the closest point in the pattern space. The closeness is measured using the distance between the two points. It is a lazy learner algorithm because it requires highest storage capacity and efficient indexing technique.

**D. Regression Techniques:** Regression analysis is a type of predictive modelling technique which investigates the relationship between two variables that is dependent (target) and independent variables (predictor). Regression technique is used for forecasting and finding the casual effect relationship between the variables. It is an effective tool for modelling and analysing data. It allows us to compare the effects of variables measured on different scales, such as the effect of price changes and the number of promotional activities. These benefits help market researchers / data analysts / data scientists to eliminate and evaluate the best set of variables to be used for building predictive models. There are number of regression techniques depends upon the nature of variables and the nature of regression line to represent the relationship between variables available to make predictions as follows:

**(i) Linear Regression:** In this technique, the dependent variable is continuous, independent variable(s) can either be continuous or discrete and nature of regression line is linear. There is only one dependent variable and

one or more than one independent variables in regression representation using a best fit straight line(also known as regression line).

**(ii) Logistic regression:** Logistic regression is used to find the probability of Success and Failure. We will use this logistic regression when the dependent variable is binary (0/ 1, True/ False, Yes/ No) in nature. Since we are working here with a binomial distribution (dependent variable), we need to choose a link function called logit function which is best suited for this distribution.

**(iii) Polynomial Regression:** In a polynomial regression technique the power of independent variable is more than 1 in equation. In this regression technique, the best fit line is a curve that fits into the data points.

As discussed above there are more regression techniques for data analytics such as Stepwise regression technique, ridge regression, lasso regression, elasticnet regression.

## IV. EARLIER WORK: FRAMEWORKS FOR MOBILE ADVERTISING

Number of frameworks for mobile advertising using big data analytics are designed in earlier research work. These frameworks are designed using different tools and technologies. We are discussing some of them as follows:

**A. Framework for mobile advertising by Le Deng:** This framework is geographic location information based ad recommendation system. The system has following functions ;It tracks the geoid and profile of the users on the basis of which ads are recommended to the customers.it also track the profile of the merchants and their products .This system enbables the correlation between the users of the mobile and the product items. Predictive model is created using the training datasets by using various machine learning techniques suvh as collaborating filtering,classification and clustering.

**Technology used and solutions:** Apache spark streaming architecture is implemented Input is taken from flume,Twitter, Zeromtqte. It chops the data into RDDs(Resilient distributed datasets) The architecture of spark streaming is given below:

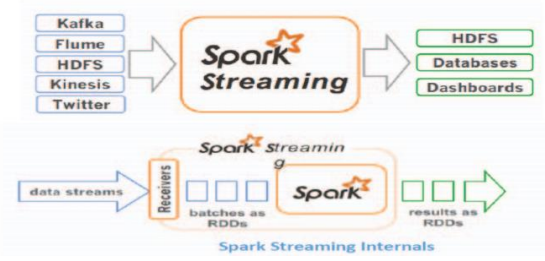


Figure 6: Spark streaming architecture

These rdds operation are applied on them after that map reduce is performed on each batch. Various datasets required are geographic information, application profile, merchant profiles and user profiles. Decision based algorithms are used because it require minimum domain knowledge .k means clustering algorithm is used where k items are used for center of k clusters. This system uses similarity analysis which construct utility matrix to recommend a new product to customer based preferences and rating. zAn in memory cluster computing platform is used in system.

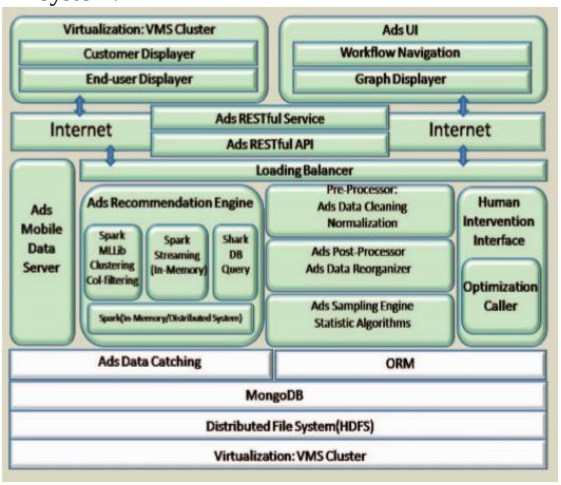


Figure 7: System component architecture[1]

Fundamental data component required are application profile, merchant profile and user profile for targeting mapping load balancer for balancing spike workloads. RBAC security is used in this system.

**B. Framework based on social network big data[9]:** In this framework the major input sources are social media such as twitter Yelp, facebook and existing advertising publisher such as yahoo and putana. MongoDB is used in this system. The architecture of the system is given below:


Figure 8:[9]

The system plays two roles one for administrator second is advertiser. Four major components for advertising are ad planning, data configuration, ad management and ad feedback. Advertisement effective analysis can be done in two ways i.e. machine learning sentiment design which consist classification and clustering technique and other is

untrained sentiment analysis design different keyboardset to decide the which sentiment is positive, negative or neutral. The following most common components will be used when using Weka to implement related machine learning algorithms.

- Instances: This is the input data.
- Filter: This is used to preprocess the data.
- Classifier/Clusterer: This is the classifying and clustering algorithm used process data.
- Evaluating: This is used to evaluate the algorithm effectiveness.
- Attribute selection: This is used to remove irrelevant attributes from the input data.

The recommendation algorithms are of their types synthesis index strategy, single index without strategy and single index with weight strategy. In recommendation strategy various variables are added in the profile with different weight points it fetches demographic baseline information after that particular algorithm is chosen to apply. The radar chart is used to define the advertisement strategies. The radar chart is spider like diagram in which each arc represents one advertising dimensional feature.

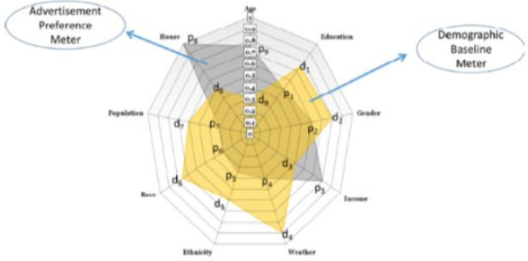


Figure 9: Advertisment spider diagram

**C. Framework based on telecom big data [10]:** This system is based on the data collected by the telecom operators to improve the accuracy of targeting the particular customer for mobile advertising, Programmatic advertising approach is used. It consists of components as follows: advertiser, DSP (demand side platform), Ad exchange and publisher. this whole system architecture consist of two parts, operators' DMP part and the mobile advertising part both are connected by API .User portrait is the basis of the system because it consist all the features of the user such as attribute, consumption ,location services. The ads are pushed to the user according to the user portrait. It classify the user as positive or negative according to the probability score. This classification models are made of two parts first is offline part which uses historic data and the other is online which uses robust and trained stable model. Up-sampling methods are used to create balanced training dataset. For model training we apply L1 constrained logistic regression algorithm. For model evaluation whether the classification model is appropriate for us or not. It uses lift chart .lift chart is used to find the lift gain by the advertiser by using data mode copare to the random user selling.

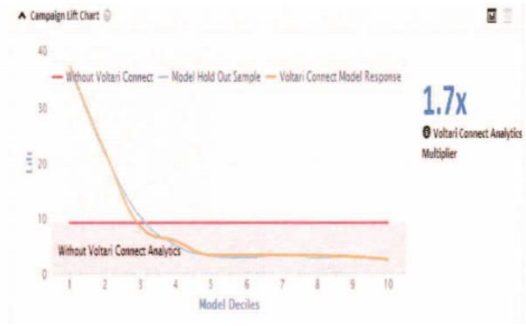


Figure 10: Generic lift diagram[10]

**D. Framework based on adaptive modelling [11]:** This Framework is an application of extreme big data for DSP (demand side platform) processing I real time auction bidding scenario for mobile display advertising[11]. This framework uses the propensity model and bid optimization strategies. It has four components: Data collection, Data Processing, Predictive Modelling and Real time scoring. There are four types of data sources used in this framework. First is bid request data in which when user accesses the publisher's website the information generated. Second is the customer data which the advertiser or publisher owns. This contains the customer purchase history and browsing behaviour as well as customer segments, usage and expenditures. Third is data about customer in the DSP ecosystem over multiple campaigns as well as bidding histories. Fourth source is external databases such as Axcioms, Experian, Neilsen and social media. This ecosystem uses ETL and data processing tasks are performed using Kafka backend services in conjuction with Spark. The predictive modelling uses two forms of processing one is Initial training model training performed by Hadoop Map Reduce services which consist if two services feature Selection and Propensity Modelling. Propensity model accepts the output of the feature selection and iterates it until all conditions of model are satisfied. The second type of processing relies on Kafka and Spark implementation. Kafka and Spark provides a near real time approach for performing adaptive modelling. The main focus of this component is to tune the scorecard until the next version of the training data scorecard arrives and begins the process all over again. The last component is real time scoring which is responsible for identifying the device, collecting the device profile and scoring the user profile against potentially hundreds of campaign scorecards within a 40 to 100 milliseconds timeframe and for computing the bid price once the user probabilities have been computed. Customer scoring is calculated from customer attributes, bid price computations. The bid determination contains two step process as follows:

1. Computation of initial bid price depending upon constraints of campaign and real time analytics of the bidding stream to detect possible patterns in the bidding time series.
2. Adjust the initial bid price to reflect the customer score.

In this framework model evaluation is done by using lift chart. The important feature of model evaluation is to determine the overall campaign model effectiveness.

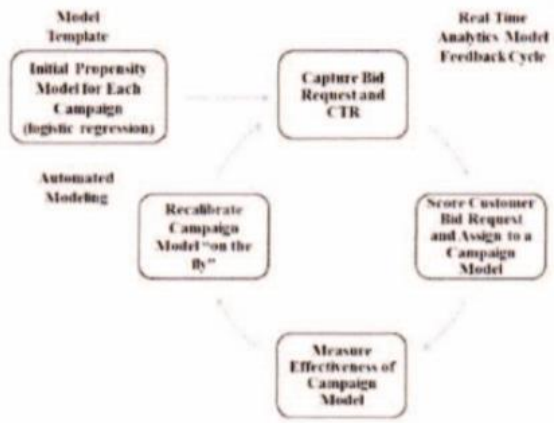


Figure 11: Real time analytics lifecycle automated modelling [11]

The complete working of the model is described in figure above: Huge number of models and model instances are stored for future references. The adaptive modelling has following features; Automated modelling, Continuous model feedback loops, Real Time streaming analytics, Huge number of models and model instances are stored for future refernces.

**E. Enabling Privacy-Aware Location based Advertising [12]:** Most of today's Location based Advertising track users personal and private information in order to be able to serve the most relevant ads to the users that raised many concerns about privacy violation. In [12], they propose PrAd, a location-based advertising model that appreciates users' location privacy. This system consist of two components one is SC (Secure Coprocessor) and the other is mPrAd. As stated in "SC can implement by using secure cryptographic processors or the recently proposed Software Guard Extensions (SGX) from Intel" .Every request is routed through SC and then goes to LBAS. The second component mPrAd collects location information from GPS sensor and performs the space encoding to get a corresponding location index i. After that mPrad creates a secure channel to SC and submits the request via that secure channel. The Secure Coprocessor receives the request, performs a private ads retrieval by using a protocol given in[14]. After that SC returns the ads to mPrad. and mPrAd forwards them to the requesting applications. Client applictions has to report to LBAS multiple times each advertisement is clicked or displayed for appropriate billing and accounting. Every time an advertisement is displayed on the application's GUI, it sends an acknowledgement to mPrAd. Based on these acknowledgements, mPrAd keeps track of how many times each advertisement is displayed. At the end of a billing period, mPrAd constructs billing vectors and reports them anonymously to the LBAS. This system uses the homomorphic encryption and k- anonymity to maintain the anonymity of ads reports [12]. This system uses the yelp dataset to evaluate the model. This privacy aware mechanism will arguably encourage more consumers to join the system and thus give rise for the market.

## V. CONCLUSION

This paper is a survey of various big data analytics techniques and tools. As there are many tools having their own features it is necessary to choose a particular technique to solve the problem. Many frameworks are developed in recent years to solve the problem of tageting the user for ad recommendation. Mobile advertising is on hype these days so it is mandatory to find the accurate framework for advertisement as it affects the business.

**Comparison table of various frameworks**

| | **Lei Deng and Jerry Gao and Chandrasekar Vuppalapati** | **Lei Deng and Jerry Gao** | **Tao Zhang, Xinzhou Cheng, Mingqiang Yuan, Lexi Xu, Chen Cheng, kun Chao** | **Donald Kridel, Daniel Dolk and David Castillo** | **Hung Dang, Ee-Chien Chang** |
|---|---|---|---|---|---|
| **Platfom used** | Spark Streaming architecture | Zarkov Mapreduce | Number of data processing methods | Hadoop map reduce, Spark and Kafka implemebtation | No of data processing tools |
| **Input sources** | Kafka, Flume, HDFS, Twitter | Yelp, Twitter, Facebook | Telecom operator database | bid request data, customer data, customer purchase history and browsing behaviour,External database | Yelp |
| **Algorithms used** | Decision making algorithm, clustering, Similarity analysis | Machine learning(classification and clustering) Untrained sentiment analysis design | ML algorithm L-1 constrained logistic regression | Adaptive modelling, Bid optimuzation | Private ad reterival algorithm.secure coprrocessper |
| **Database used** | MongoDB | MongoDB, In memory | MPP database | Axcioms, Experian, Neilsen and social media | |

| | | | | | |
|---|---|---|---|---|---|
| **Data components** | Application Profile, merchant profile, user portrait | User profile, demographic information | User profile and Network used | | |
| **Recommen-dation strategy** | It consist of collaborative and content based filtering | Synthesis Index strategy, single index without weight strstegy, single index with weight strategy | | | |
| **Model evaluation** | Lift chart | Radar chart | Lift Chart | Lift chart | Unit square size |

## VI. REFERENCES


[1]. Lei Deng and Jerry Gao and Chandrasekar Vuppalapati," Building a Big Data Analytics Service Framework for Mobile Advertising and Marketing" IEEE First International Conference on Big Data Computing Service and Applications, 2015.

[2]. Nader Mohamed and Jameela Al-Jaroodi," Real-Time Big Data Analytics: Applications and Challenges ",IEEE ,2014.

[3]. YojnaArora and DrDineshGoyal," BigData: A Review ofAnalytics Methods & Techniques",IEEE,2016.

[4]. Babak Yadranjiaghdam, Nathan Pool, Nasseh Tabrizi," A Survey on Real-time Big Data Analytics: Applications and Tools ", International Conference on Computational Science and Computational Intelligence,2016.

[5]. Mrunal Sogodekar, Shikha Pandey, Isha Tupkari and Amit Manekar ," BIG DATA ANALYTICS: HADOOP AND TOOLS", IEEE Bombay Section Symposium (IBSS),2016.

[6]. Yavuz CDQED and Seref Sdgourgox," Big Data Anonymization with Spark",IEEE,2017.

[7]. Wenjie Yang, Xingang Liu* and Lan Zhang , Laurence T. Yang ," Big Data Real-time Processing Based on Storm , 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications,2013.

[8]. Poonam vashisht and Vishal Gupta," Big Data Analytics Techniques: A Survey",IEEE,2015.

[9]. Lei Deng and Jerry Gao," An Advertising Analytics Framework Using Social Network Big Data", 5th International Conference on Information Science and Technology (IClST) April 24--26, 2015, Changsha, Hunan, China.

[10]. Tao Zhang, Xinzhou Cheng, Mingqiang Yuan, Lexi Xu, Chen Cheng, kun Chao," Mining Target Users for Mobile Advertising Based on Telecom Big Data",IEEE,2016.

[11]. Donald Kridel, Daniel Dolk and David Castillo,"Adaptive modelling for real time analytics: The Case of big data in mobile advertising" , Hawaii International conference on system science,2014.

[12]. Hung Dang, Ee-Chien Chang," PrAd: Enabling Privacy-Aware Location based Advertising",ACM,2015


"


Kawaljit Kaur received the B.Tech degree in Computer Science and Engineering from Khalsa college of engineering and technology Amritsar(Punjab), INDIA in 2016 and pursuing M.Tech degree from Guru Nanak Dev University Amritsar(Punjab), INDIA. Her area of interest is Big Data analytics.

Sandeep Sharma received the B.E. in Computer Engineering from Pune University, INDIA in 1994,M.E. degree in Computer Engineering from Thapar institute of Engineering and Technology Patiala, INDIA in 2000 and Ph.D. degree from Guru Nanak Dev University Amritsar(Punjab), INDIA in 2011.He is working as Head of Department in department of Computer Engineering and Technology and he is also acting as chief security information officer of the university. His area of interest is Parallel Processing, Multistage interconnection network, Fault Tolerance and cloud computing.