Big Data Analytics Service Framework for Advertising and Marketing

[1] Mr. Ganapati V Ourasang, [2] Mr. Prakash V Parande

Student, M. Tech, Dept. of CSE, VTU- CPGS, Visvesvaraya Institute of Advance Technology, Muddenahalli, Bengaluru, Karnataka, India (ourasang.ganesh@gmail.com)

Assistant Professor, Dept. of MCA, VTU-CPGS, Visvesvaraya Institute of Advance Technology, Muddenahalli, Bengaluru, Karnataka, India (Prakashvp2010@gmail.com)

Abstract

Enormous growth in adopting digital marketing this leaves a large storage of data to users. Daily updating data makes conduct and mentioning to operators makes more number of selections in related to marketing and advertising. This needs an up to date updation to clients and application dataset. Customer is busy with the more volume of data for own advertisement or users concern to check nearby services which are the services already update to dataset. If the customer wants to add own service, go throw login and add services. For both these consumer should need username and password. Customer interest on digital marketing, it makes customer to wait somewhat little more time and storage space is also increased. Today there is go on increasing in the craze towards digital marketing. More number of people increase with respect to storage space and increase in ad. Due to increase in large volume of data and users advertising to obtain a huge marketing network, this leads to increase in service time. To avoid this problem we are going to use MapReduce framework.

Key Words: Big Data, HDFS File directories, MapReduce, Hadoop.

I. INTRODUCTION

There is a possibility of growth in profit and customers experience, this gives marketers and advertisers come in real-time customer engagement to prove brand loyalty. An advantage is a daily grow in ad day by day and along with this marketing. Big facts are a variety of huge facts, so big facts applications are large in size. Present proposed framework that supports an offline publicizing operations in which the selected analytics performances are used to give advertising that supports based on collected big documents on users profiles, access actions. A design solution that gives which are the nearby welling service provides are available. By this service methods, user can select his interest one suppose that promotion or advertising services may be schools, banks, hospitals and hotels.

An advertiser may store his advertising to big data applications just by entering advertising details with the login ID and password then only we will enter through a login then signup if user is not belonging to that login it is not possible to interact with an application. The advertising gives three different possibilities or challenges huge growth adopting digital marketing. Customer is busy with his large amount of datasets of advertising and marketing.

Customers interest on digital marketing. Adopting digital marketing it is increasing day by day. Every day almost 1 million users are to be going to had craze in the mobile advertising and marketing. Customers busy with his great quantity of dataset, he is try to load the own advertise to marketing purpose or search for nearby place which is to be best choose from an existing details. There is a possibility of the data can upload by users own advertise for marketing purpose and search for anything need an advertised data for own use. Customer interest in digital marketing is also increases a huge volume of dataset and number of users to the digital marketing.

A 61% of the USA subscriber maintained a smartphones. The mobile devices come with various form factors, technologies, data points, and operating systems. As the difficulty of mobile phone increases and as the mobile users demographics and personal preferences differs the linked datasets with the units and users will increase intensely. Market Analysis discloses that mobility increasingly defines digital marketing. The social and commerce activities that consumers busy with their handsets are filtering mobile adverting. Hadoop definitely attracts much care as it is the first open source distributed computing environment. But there are other platforms that have interesting advantages to the typical in its implementation, especially in the real-time analytics of dynamic information where hadoop does not meet the requirement. Storm is a stream processing framework and focuses on continuous computation.

The method is labor intensive when given large training sets, and did not add popularity until the 1960`s when amplified computing power became available. It has since been widely used in the area of

pattern recognition. The recommendation process is performed in three steps each of which is handled by a separate component. Admin-This admin part covers user detailed history. Server- In this all the dataset as in form of HDFS file directories stored with the JSON database. User- In this user wish to login and check for the nearby place service. Real-time advertisement, real-time customer engagement are most current information available to the server at the online mode. Offline processing is typically batch-oriented and often includes large volumes of data being handled with little or no intrusion. Facts can be nursed to streaming through distributed messaging queues like flume, twitter or plain old TCP sockets.

A. Statement of problem

- By entering latitude and longitude of a place it gives a nearby place services like Hospitals, schools, hotels.
- Using of a same password and username for all services.
- This project provides an offline mode of execution

B. Objective

Using location latitude and longitude of range limit only it displays nearby place services like Hotels, Hospitals, collages name.

II. BACKGROUND

The previous items or products are already deal with the clustering. A clustering means a more number of machines or tools can share a large volume of data for dealing with the big data applications. The matching with input data by decision tree algorithm and resulting facts from a clustering of information to produce a resulting form, in customers required data.

An existing system that gives matching of facts what is entered by client it matches then it help to customer access easily. A clustering of previous research is helps to find a reduce form to our project and display a result in the reduced form. The same category of data is grouped at one place or this referred to as clustering. It provides easy access for the threshold values input by user to fetch within this range.

Disadvantage of existing system

Large volume of facts matching each time with decision tree algorithm makes more complexity in building an application.

Grouping of same matching data in a separate nodes or system with huge volume of information by clustering is time consuming and cost also increased. Because if more a data need more clusters to separate matching facts.

K-Nearest Nieghbour don't construct the classifier until they receive unknown sample, the classification is slower because of learner specification delay.

III.   LITERATURE SURVEY

James Manyika et al., " Disruptive technologies: Advances that will transform life, business and the global economy", May 2013. [3], this gives alternative techniques to entrepreneurs with technologies or methods to a market delivering. No one change the product trading ways. Source to a many number of different terms to improve and transform life trade and world economy. A drawback is, its needs in-depth analysis of key potential is difficult. Try to adopt technologies may take years of time for different transforming methods.

Giamas Alex, "Spark, Storm and Real Time Analytics", 2014. [6], these are also aims like hadoop for the computation of distributed events. Distributed talks capabilities of computation are to be done with spark and storm as like hadoop. The advantages are its helps to computation of distributed computation capabilities. It is applied for large scale of data for processing written in scale presenting using Apache spark. For batch processing STORM- YARN is used. Disadvantages are, all difficulties with hadoop are not overcome with adapting of storm and spark.

Martin Kihn and Mike McGuire, "Gartner Webinars, Mobile Marketing and Data-Driven Marketing", Research VP, 14 May 2014 [7], as like a hadoop, spark is also an open source cluster calculating environment. Optimized interactive workloads for queries of a spark enabled in memory distributed datasets. The advantages are, unlike hadoop, integration between distributed data and manipulation is done with scale. Side by side over the hadoop file system for complementary of jobs distribution that support datasets by a third party clustering framework. The disadvantages are, a complementary of jobs distribution of a hadoop file systems for side by side support needs a third party clustering framework. It cashed in memory to reduce their latency of access for cluster computing framework to these types of workloads.

Jiawei Han and Micheline Kamber and Jian Pei, Morgan Kaufmann, "Data Mining: Concepts and Techniques", Elsevier Inc. (2011) [8], how a data mining is to be done with large capacity of facts for security techniques. Now a days, a present trend is getting an intelligence report from blog, twitter etc. The patterns of data for security purpose also discussed in this paper. The advantages are, it

provides security in networks for huge capacity of data. In this, patterns to the data for matching. The disadvantages are, more data need more confidential security, but by just patterns it is difficult.

Pasquale Lops, et al., "Content-based Recommender Systems: State of the Art and Trends [9], directing to user in a personalized way to intersecting objects in a large space of possible options. It is depends on a user profile in which preferences of interests are to be stored. The advantages are, easily finding to a user interest of recommend depends on their profile. The design of basic architecture makes to know users interest recommend. The disadvantages are, high level architecture makes a difficulty to sudden understanding to user interest recommend.

Real-Time, Online and Offline Complex Event Processing",[10] Allows to us training data is to be online or offline for complex object processing. For large collection of data handling of objects may deal into the complex process.

X Lieu, "Knowing of Big Data Processing and analytics", [11].First of all knowing of the processing and analytics of big data is important. In this we learned that how actually big data processing and analysis is done.

T Das, Spark Streaming,[12] Informs how data streaming is done with the help of a spark streaming, this allows collecting information online throw twitter, Facebooksocial networks. What base of idea customers can have at present?

IV. PROPOSED SOLUTION

The presented application is to be developed to avoid the disadvantages with an existing system. An existing system that provides decision tree algorithm but avoiding that we choose a new matching concept called mapping. User input is stored first then, related to that input mapping is necessary. A result is produced in the reduce form called the mapreduce framework. This framework finally displays a result of reduced form which are required by nearby point ordering form one to last. A first reduce displays data is most convenient to customer. Any how choosing the data is depends on customer.

Advantages of proposed system

The developers do not have the idea, where the data is distributed among nodes. The programmer just deals with the mapping and reducer framework.

Map and reduce functions are done easily because user input stored in node and then according to that input mapreduce framework is to be processed.

This framework that provides a login process with cloudera to begin with a big data application.

Add on service is activated, if we need to add the any advertisements to the cloud. It depends on the interest of customer already stored data accessed and helpful to access what we need. Satisfaction to customer is done with mapreduce framework.

An add on services that gives entering the latitude and longitude that's depending on these values with in an area customer required nearby point advertisements are to be displayed in the reduced form.

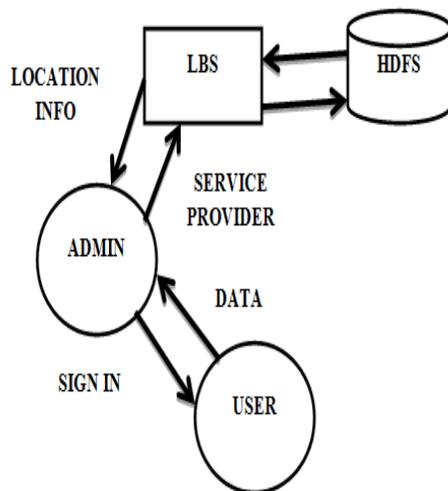The following figure shows the represent flow of a sequence process.

Fig 1 System Architecture

The fig 1 shows system architecture, in our system architecture for location based services to user is to sign in for authentication, after authentication of the administrator is able to access a location based service. The service provider that sends details to a LBS that regarding a reduced form location based service are nearby places, these are already stored in the form of file data as in the HDFS. It provides all details of LBS to service providers as to administrator to user. Finally an user gets information about nearby places as send by a service providers from a location based services which is to be stored in the HDFS for user to access any time wants the details of nearby places.

A. SERVICE PROVIDER MODULE

This module is used to add different services to HDFC directory. Adding a location name, users all details for login and sign up, sign in also. Service provider module provides if changes need to a present data it allows add services option.

B. ADMIN MODULE

In this interaction between user and service provider takes place. It checks user and then only allows go for next flow of service to user. If unauthorized person wants to deal with an application it will not allow.

C. USER MODULE

This module we enter a user related all data. If user wants to access a required result or a location details a user has to enter location name and if he wants to enter location name and if he want to login need to enter correct username and password

V. CONCLUSION

This paper represents a novel approach to providing a location based ad recommendation system using the present state of art machineries. The paper provides a decision based approach to work for innumerable use cases associated with pushing relevant ads towards the clients. The objective of the paper is to endure the whole process of complete testing and wait for result which would enable us to put forward a scalable big data ad processing platform in the present market. It also provides a flier data analytics approach for the clients to view their end users. Once an experimentation of this approach is beta-tested based on the reviews of the client, we should improve the data analytics component our system.

The range of the current paper is very large. A lot of features can be added to the system to reach high scalability supporting real-time processing and recommendation. Currently our system that handles only offline modeling and training. In the future, we plan to provide online modeling and training of datasets to enrich user experience.

REFERENCES

[1] Google, "Android, the world's most popular mobile platform", 11 Jan 2015.

[2] Chunk Jones, "Why 10 Million iPhones Means A Lot More Than 9 Million", Sept 2013.

[3] James Manyika et al., " Disruptive technologies: Advances that will transform life, business and the global economy", May 2013.

[4] Cohen H.." How Your Audience Uses Mobile Now", 2013.

[5] Martin Kihn and Mike McGuire, "Gartner Webinars, Mobile Marketing and Data-Driven Marketing", Research VP, 14 May 2014.

[6] Giamas Alex, "Spark, Storm and Real Time Analytics", 2014.

[7] Jones M. T. "Spark, an alternative for fast data Analytics", 2012.

[8] Jiawei Han and Micheline Kamber and Jian Pei, Morgan Kaufmann, "Data Mining: Concepts and Techniques", Elsevier Inc. (2011).

[9] Pasquale Lops, et al., "Content-based Recommender Systems: State of the Art and Trends.

[10] Real-Time, Online and Offline Complex Event Processing", Nov. 5, 2014.

[11] X. Liu, "Understanding Big Data Processing and Analytics", 2013, sept.19.

[12] T. Das Spark Streaming.