

Application of Map Reducing Techniques in Online Food Marketing: An Algorithmic Approach

¹Prashant Bhat, ²Prajna Hegde,

¹Assistant Professor
Department of Computer Science, Garden City
University, Bengaluru
prashantrcu@gmail.com

²Research Scholar
Department of Computer Science, Garden City University, Bengaluru
Prajna_rh@yahoo.co.in

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 16 April 2021

ABSTRACT: Food industry is dynamic industry as it depends highly on consumer interest and timing. Big Data Analytics can bring considerable revolution in food industry. It can be used to improve operational efficiency, timely delivery and also it can be used to personalize customer experience. In the field of marketing, Big Data Analytics helps in deciding pricing, demand forecasting as well as product promotion. The objective here is to show how Map Reduce technique can be used to process and analyze Big Data to extract useful information. This paper gives insight into use of Big Data Analytics in food marketing by analyzing huge amount of data using Big Data processing tool Apache Hadoop.

Keywords—Big Data; Big Data Analytics; Hadoop; Map Reduce;

I. INTRODUCTION

Digital platforms provide handful of opportunities in the field of business. Consumers are linked through mobiles and networks. Customer purchase details, web clicks and other activities online are collected and stored. These can be analyzed to get information which help company to target customers for marketing. This helps to find the patterns and to detect the factors which leads to customer purchase decisions.

Big Data means big opportunity. With the help of Big Data, data is processed and analyzed in real time within short time. As a result quick action can be taken which is difficult to achieve using any other technology. Solutions generated from Big Data Analytics gives new approaches to solve some of the key marketing essentials and gives impressive outcomes. These outcomes can improve the execution of marketing functions. Here in this paper an attempt is made to showcase how Big Data Analytics can be used to retrieve useful data and what difference it can create in the industry. Along with this, the paper also gives an algorithm which processes Big Data using Map Reduce technique to retrieve useful information.

II. RELATED WORK

Authors Dandan Tao, Pengkun Yang, Hao Feng say Big Data has the capacity to convert large amount of data into information that helps in taking business decisions. As a result data mining techniques are used in the field of supply chains in the food industry. Digital text data in food industry has got importance because of advanced Big Data Analytics. Here Authors mentioned word level analytics, text classification, text clustering, topic modeling etc. Food safety, food safety, food fraud surveillance, dietary pattern characterization, consumer opinion mining using text data analysis have been illustrated. The intention of this paper is to give highlight for intelligent decision making to improve food industry including food production safety and human nutrition.

Authors Abdul Jabbar, Perviaz Akhtar, and Samir Dani says Data in huge amount are collected from different resources which has created many opportunities for B2B or in marketing field. This is a challenge for marketing organization as online display advertising, cleansing and processing of data in huge volume is complicated. Also it is difficult to take real time decision and comparative implications. This paper makes use of problematization approach. Here authors have given relation between Big Data, and programmatic marketing, real time processing and decision making for marketing organization. This paper gives direction to the future research. The paper also builds interdisciplinary dialogues that covers engineering frameworks like Apache storm and Hadoop.

Authors Sachin Thanekar, K Subrahmanyam, and A.B Bhagwan presented a paper which says huge volume, high velocity and veracity are the sign of Big Data. As a result of these characteristics of Big Data, there are many difficulties in processing it. Using traditional database or systems it is difficult to store, analyze and process Big Data because data are structured, semi-structured, unstructured. Data protection, capture, searching, storage and visualization are the main challenges when it comes to Big Data. Map Reduce processes huge data in parallel way,

and in reliable manner. This paper focus on Map Reduce, its challenges and opportunities, recent trends which can help researchers in future.

Authors Shweta Pandey, Vrinda Tokekar presented a paper in which they said there are lot of advantages, better service delivery with Big Data in business. Big Data is creating positive change in business organizations by helping in better decision making. Big Data not only has advantages but also has several issues, challenges. Definition of Big Data explains the complexity due to large volume of data which makes it difficult to process using existing system. In this paper authors explained Big Data, role of Map Reduce in Big Data. MapReduce is extendable and efficient when compared to other models.

Authors Mohammad Saleem Alshura, Abdelrahim.M Zabadi, Mohamad Abughazaleh presented a paper in which authors tried to present integrated view of Big Data in marketing field based on literature review. Systematic review is conducted by the authors to study opportunities, benefits of Big Data. There are lots of things to be studied in the field of Big Data in order to reach marketer's goal. Big Data in marketing is still in initial stage. The paper is related to development of marketing activities on customer basis. In this paper author's detailed study gives clear view of opportunities and challenges of Big Data which can help researchers in upcoming days.

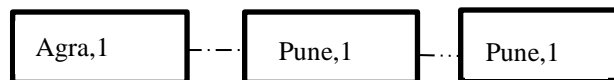
III. PROPOSED WORK

Customer involvement or how customer see and interact with the product is the main factor for marketing. Product can be improvised with the help of business intelligence provided by Big Data Analytics. Big Data can be used to find the factor which effects more on customer which makes customer come back again. By collecting customer data, activity of customer online and change in trend can be detected. It makes marketing more sophisticated. Big Data provides 360 degree view using which customer oriented content can be developed.

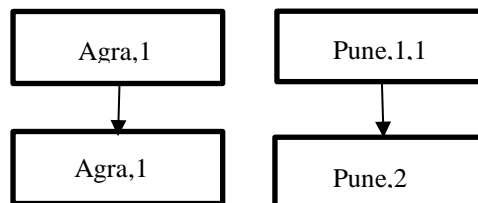
In this paper an attempt is made to show how data collected from Zomato which has details about restaurant and other details like customer rating, photo count and details of city etc., can be used to extract information using Map Reduce technique.

The Map Reduce technique in this novel approach works as follows.

- Mapper reads each row and splits it into tokens. Assign value 1 to each City name.



- The values are shuffled and reducer combines the values with same key.



The following algorithm shows how the dataset of Zomato can be processes and analyzed to extract some useful information

Step1: Write a program restaurants in java which does the following task

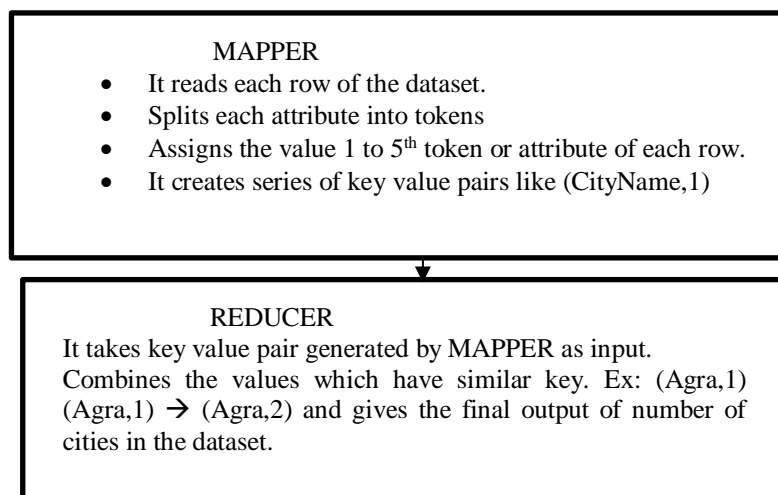


Fig.1 Map Reduce program execution in Hadoop

Step2:

Create a folder called 'Zomato' and store the written java program and dataset in a folder called input. Then grant all the access permission to 'Zomato' folder.

Step3:

Start all nodes of Hadoop by typing start-all.sh on command prompt. Then, Type following command on command prompt

```
export HADOOP_CLASSPATH=$(hadoop classpath)
Echo $ HADOOP_CLASSPATH
```

Step4:

Store input folder in HDFS

Hadoop fs -copyFromLocal <input folder address> /user/Hadoop
i.e hadoop fs -copyFromLocal /home/hadoopusr/zomato/input /user/Hadoop

Step5:

Create a folder class inside zomato folder.

javac -classpath \${HADOOP_CLASSPATH} -d <address of class folder> <restaurants program address>
i.e javac -classpath \${HADOOP_CLASSPATH} -d /home/hadoopusr/zomato/class
/home/hadoopusr/zomato/restaurants.java

Step6:

Create jar file.

Jar -cvf restdata.jar -C class/.

Step7:

Run the program

hadoop jar <jar file address> <name of class inside program> <input data address in HDFS> < address of output to be stored>

I.e. hadoop jar /home/hadoopusr/zomato/restdata.jar restaurants /user/Hadoop/input /user/Hadoop/output (Fig.1)

Step8:

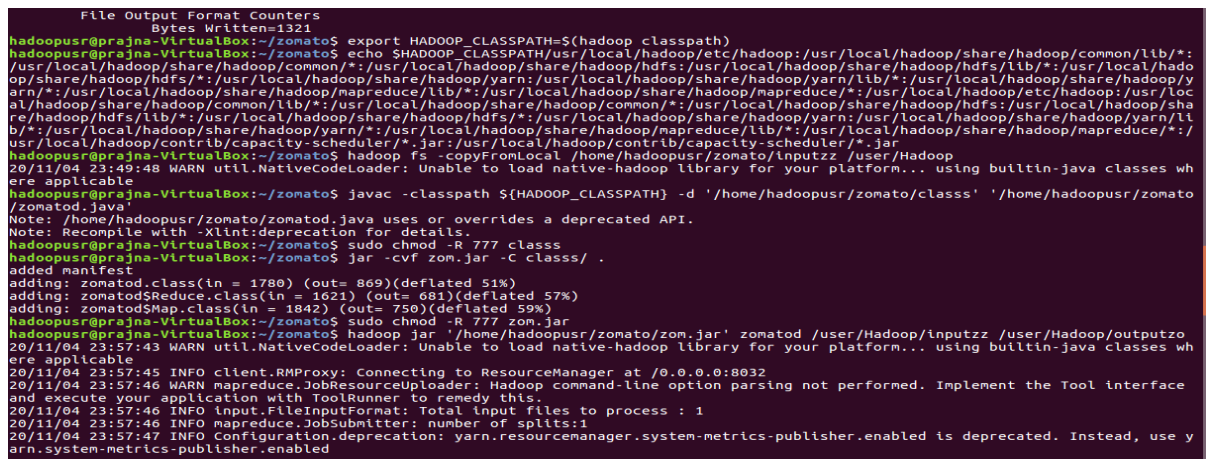
See the output.

hadoop fs -cat <output folder address>

i.e hadoop fs -cat /user/Hadoop?output//part-r-000000

Or one can also see the output by visiting localhost -50070 from browser.

The above novel approach gives the city names along with the number restaurants in that particular city. This information is extracted by processing huge dataset of zomato. This novel approach makes use of attribute city name. This work is a best example for how Map Reduce technique can be used to extract information which can be used by the organization. Here by knowing the number of restaurants in every city, the city with minimum



```
File Output Format Counters
  Bytes Written:1321
hadoopusr@prajna-VirtualBox:~/zomato$ export HADOOP_CLASSPATH=$(hadoop classpath)
hadoopusr@prajna-VirtualBox:~/zomato$ echo $HADOOP_CLASSPATH/usr/local/hadoop/etc/hadoop:/usr/local/hadoop/share/hadoop/common/lib/*:/usr/local/hadoop/share/hadoop/common/*:/usr/local/hadoop/share/hadoop/hdfs:/usr/local/hadoop/share/hadoop/hdfs/lib/*:/usr/local/hadoop/share/hadoop/yarn/lib/*:/usr/local/hadoop/share/hadoop/yarn/*:/usr/local/hadoop/share/hadoop/mapreduce/lib/*:/usr/local/hadoop/share/hadoop/mapreduce/*:/usr/local/hadoop/share/hadoop/common/lib/*:/usr/local/hadoop/share/hadoop/common/*:/usr/local/hadoop/share/hadoop/hdfs:/usr/local/hadoop/share/hadoop/hdfs/lib/*:/usr/local/hadoop/share/hadoop/hdfs/*:/usr/local/hadoop/share/hadoop/yarn:/usr/local/hadoop/share/hadoop/yarn/lib/*:/usr/local/hadoop/share/hadoop/yarn/*:/usr/local/hadoop/share/hadoop/mapreduce/lib/*:/usr/local/hadoop/share/hadoop/mapreduce/*:/usr/local/hadoop/contrib/capacity-scheduler/*:jar:/usr/local/hadoop/contrib/capacity-scheduler/*:jar
hadoopusr@prajna-VirtualBox:~/zomato$ hadoop fs -copyFromLocal /home/hadoopusr/zomato/inputzz /user/Hadoop
20/11/04 23:49:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoopusr@prajna-VirtualBox:~/zomato$ javac -classpath ${HADOOP_CLASSPATH} -d '/home/hadoopusr/zomato/class' '/home/hadoopusr/zomato/zomatod.java'
Note: /home/hadoopusr/zomato/zomatod.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
hadoopusr@prajna-VirtualBox:~/zomato$ sudo chmod -R 777 class
hadoopusr@prajna-VirtualBox:~/zomato$ jar -cvf zom.jar -C class/ .
added manifest
adding: zomatod.class(in = 1780) (out= 869)(deflated 51%)
adding: zomatod$Reduce.class(in = 1621) (out= 681)(deflated 57%)
adding: zomatod$Map.class(in = 1842) (out= 750)(deflated 59%)
hadoopusr@prajna-VirtualBox:~/zomato$ sudo chmod -R 777 zom.jar
hadoopusr@prajna-VirtualBox:~/zomato$ hadoop jar '/home/hadoopusr/zomato/zom.jar' zomatod /user/Hadoop/inputzz /user/Hadoop/outputz
20/11/04 23:57:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
20/11/04 23:57:45 INFO client.RMProxy: Connecting to ResourceManager at 0.0.0.0:8032
20/11/04 23:57:46 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
20/11/04 23:57:46 INFO input.FileInputFormat: Total input files to process : 1
20/11/04 23:57:46 INFO mapreduce.JobSubmitter: number of splits:1
20/11/04 23:57:47 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
```

Fig.2 Output showing number of restaurants in different cities

number of restaurants which provide zomato service can be known (Fig.2). Same way, the city with maximum number of restaurants having zomato service can also be found.

Big Data Analytics allows to extract information as shown in the above novel approach. Marketing can be optimized using Big Data Analytics. Specific audience can be targeted based on purchase history, email address or customer web search. Costs, pricing and statistics of a company can be analyzed to make company's performance more reliable.

IV. CONCLUSION

Using Big Data Analytics in marketing not only saves money but also efficiency. Big Data in customer relationship management combines organizational data and the data that resides outside the system. Big Data Analysis results in finding patterns in data, and trends which increases sales opportunities. Critical behavioral understanding which requires some action to return customer base can be obtained using Big Data. In this paper an attempt has been made to show how large amount of data can be analyzed using tool like Hadoop, Map Reducing technique and information can be retrieved. The retrieved information can be utilized so as to improve the working efficiency the organization.

REFERENCES

1. Dandan Tao, Pengkun Yang, Hao Feng, 2020, "Utilization of text mining as a big data analysis tool for food science and nutrition", *Comprehensive reviews in food science and food safety*, 19 2, 875-894 .
2. Abdul Jabbar, Pervaiz Akhtar, Samir Dani, "Real-time big data processing for instantaneous marketing decisions: A problematization approach", *Industrial Marketing Management*, Volume 90, 2020, Pages 558-569, ISSN 0019-8501
3. Sachin Thanekar, Subrahmanyam, K, Bagwan, A, 2016, "A Study on MapReduce: Challenges and Trends", *Indonesian Journal of Electrical Engineering and Computer Science* 4. 176. 10.11591, pp176-183.
4. Alshura, Mohammed, Zabadi, Abdelrahim, Abughazaleh, Mohamad, 2018, "Big Data in Marketing Arena. Big Opportunity, Big Challenge, and Research Trends: An Integrated View", *Management and economics review*, 3. 10.24818/mer/2018.06-06.
5. Pandey, Shweta, Tokekar, Vrinda, 2014, "Prominence of MapReduce in Big Data Processing", *Proceedings 2014 4th International Conference on Communication Systems and Network Technologies*, CSNT 555-560.
6. Khezr, Seyednima, Navimipour, Nima, 2017, "MapReduce and Its Applications, Challenges, and Architecture: a Comprehensive Review and Directions for Future Research", *Journal of Grid Computing*. 15. 1-27. 10.1007/s10723-017-9408-0.
7. Ramos, Marcelo, Tasinaffo, Paulo, Almeida, Eugenio, Achite, Luis Cunha, Adilson, Dias, Luiz, 2016, "Distributed Systems Performance for Big Data", 10.1007/978-3-319-32467-8_64.
8. Elgendy, Nada, Elragal, Ahmed, 2014, "Big Data Analytics: A Literature Review Paper", *Lecture Notes in Computer Science*, 8557. 214-227. 10.1007/978-3-319-08976-8_16.
9. Shu, Hong, 2016, "Big data analytics: six techniques" *Geo-spatial Information Science*, 19. 1-10. 10.1080/10095020.2016.1182307.
10. Abdelhafez, Hoda, 2016, "Mining Big Data in Telecommunications Industry: Challenges, Techniques, and Revenue Opportunity"
11. Sucharita, V, Rao, Parameswara & Kalyan, A., Rajarajeswari, P, 2018, "Big Data Mining: A Forecast to the Future", 10.4018/978-1-5225-2947-7.ch003.
12. Velusamy, Revathi, K.R, Rakshitha, .K, Sruthi, .S, Guruprasaath, 2017, "Big data with hadoop – for data management, processing and storing", *International Research Journal of Engineering and Technology*, 4. 1111-1116.
13. Alexandra Amado, Paulo Cortez, Paulo Rita, Sérgio Moro, "Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis", *European Research on Management and Business Economics*, Volume 24, Issue 1, 2018, Pages 1-7, ISSN 2444-8834
14. Maleki, Neda, Rahmani, Amir, Conti, Mauro. 2019, "MapReduce: an infrastructure review and research insights", *The Journal of Supercomputing*. 75. 1-69. 10.1007/s11227-019-02907-5.
15. Vaidya, Madhavi, 2012, "Parallel Processing of cluster by Map Reduce", *International Journal of Distributed and Parallel systems*, 3. 167-179. 10.5121/ijdps.2012.3113.