BFSMpR:A BFS Graph based Recommendation System using Map Reduce

Vivek Pandey
Dept. of Computer Science and Engineering
ShriShankaracharya Technical Campus
Junwani, Bhilai(C.G.)
vivekjdp@gmail.com

Dr. Padma Bonde
Dept. of Computer Science and Engineering
ShriShankaracharya Technical Campus
Junwani, Bhilai(C.G.)
bondepadma@gmail.com

Abstract——Nowadays, Many associations, organizations and analysts need to manage huge datasets (i.e. Terabytes or even Petabytes). A well-known information filtering algorithm for dealing with such large datasets in an effective way is Hadoop Map Reduce. These large size datasets are regularly known to as graphs by many frameworks of current intrigue (i.e. Web, informal organization). A key element of the graph based recommendation system is that they depend upon the neighbor's interest by taking minimum distance into account. Generally recent day proposal frameworks utilize complex strategy to give recommend to every user. This paper introduced an alternate approach to give suggestions to users in used of an un-weighted graph using a Hadoop iterative MapReduce approach for the execution.***)***

Keywords-*Un-Weighted Graphs, Hadoop, MapReduce, Parallel Execution, Recommendation System, Breadth first search(BFS).*

*****

I. INTRODUCTION

In Big Data environment, we will experience scalability and efficiency problems as it has to work on huge amounts of data. Many of the traditional recommender frameworks will provide same rating and ranking services to different users. So these paper introduce a Graph based service recommender system that executes on HadoopMapReduceFramework .

MapReduce is a programming model and related usage for handling and creating extensive informational collections with parallel execution and widely calculation on a bunch. MapReduce has been ideated by Google's specialist Jeffrey Dean and Sanjay Ghemawat in 2004 for dealing with huge informational collections which are normally figured by hundreds or thousands machines to complete in the most brief conceivable period [1][17]. MapReduce works through two concepts called Map() and Reduce(). The Map() work takes an info component and produced an arrangement of intermediate key-value combines and passes it to the Reduce() work. The Reduce() work takes the arrangement of the middle of the road key-value sets, combines all the transitional qualities for a specific key and delivers a littler arrangement of merged output values.

The energy of MapReduce is that it permits software engineers without a profound involvement in parallel and circulated frameworks to effectively use the asset of a vast dispersed framework, breaking a calculation into small jobs that keep running in parallel on different machines, and scales effortlessly to substantial bunches of modest product PCs. Additionally, another key advantage of MapReduce is that it smartly handles the failures and crashes of servers. Actually if a server crashes, MapReduce runs the undertaking on an alternate machine that is why hadoop is a fault tolerant framework [2].

A best execution of MapReduce is the open-source system Apache Hadoop [3]. It was created predominantly by Yahoo! in spite of the fact that is additionally utilized by different organizations, for example, Facebook, Amazon and Last.fm [4]. Hadoop is the used HDFS (Hadoop Distributed File System) file systems which was intended to store a large amount of information, so as to enhance stockpiling and get to operations to few vast record, as opposed to a ton of little document [5].

These days, numerous frameworks of current interest to mainstream researchers can helpfully be spoken to as charts [6].. Each of systems comprises of an arrangement of hubs speaking to, for example, PCs or switches on the Internet or individuals in an informal community associated each other by edges, speaking to information associations between PCs, relationships amongst individuals, etc. Frequently these graphs are comprised of an enormous measure of hubs and edges. In which the hubs are the serves and interconnection between two hubs with a wire are called edges.

We considered an alternate approach to give suggestions to users of systems. Our program expects to visit an unweighted graph beginning from an arrangement of hubs (additionally called key nodes) to discover which hubs can be come to by the key nodes inside a movable most extreme separation. It has been actualized using MapReduce keeping in mind the end goal to deal with enormous datasets.

This paper is sorted out as takes after. Segment 2 depicts related works which has been used as beginning stage for our usage of the venture. Area 3 shows the issue and research mechanism and usage. Segment 4 portrays every one of the

analyses we performed using diverse arrangements of information. At last, we finish up inspection.

II. REVIEW WORK

Discussed in [7] paper, Author proposed a new approach to improve the quality of collaborative filtering recommendation systems. The algorithm combines item clustering and weighted slope one scheme..

Paper [8] introduced two semantic social recommendation algorithms called Node-Edge-Based and Node-Based, these algorithms recommend an input item to a group of users.

[9], in this paper a combined approach of user-user collaborative filtering (CF) and item-item CF has been presented to generate recommendations on Hadoop cluster using Apache Mahout, a library for machine learning algorithms.

[10], Recommendation engines are a natural fit for analytics platforms. They involve processing large amounts of consumer data that collected online, and the results of the analysis feed real-time online applications.

Paper [11] presented combined Collaborative Filtering using Mahout on Hadoop for movie recommendation. By combining User-based and Item-based CF, accuracy of the results gets improve. Hadoop has increased throughput. Because of multiple computer nodes, time taken for solving problem has been reduced.

III. BFSMPR RECOMMENDATION SYSTEM MECHANISM

Numerous frameworks that are consistently used now can be called as graphs. Some of them are informal organizations, Internet and manyothers. A hefty portion of those frameworks utilize a technique to give recommendation to the user of the system. Some illustration are YouTube [20] that proposes you to watch videos, Facebook prescribes peoples that you could know, Amazon [21] [22] recommendssame item you purchased at certain criteria. Regularly those recommender frameworks utilize complex method to give advices to every user. The advices can be founded on estimations of likeness between products or users. Existing recommender frameworks can be sorted into two unique Techniques: content-based and synergistic sifting [19] [20].

In content based approach the framework suggest product to a particular user, by using the history of user purchases. For this situation the user history are utilized and make up a vector of components where estimation of an element can be set up by TF-IDF (Term frequency–Inverse Document Frequency) algorithm.Another user is coordinated against the database to find neighbours, which are different users who have generally had comparative taste to him/her. Products that the neighbours like are then prescribed to the user, as he/she will presumably additionally like them [21

Before beginning clarify how our calculation functions work, we need to enhance what BFS [12],[13]implies. Expansiveness initially hunt is a graph traversal calculation. The pursuit starts from the root hub and the neighboring hubs are visit until there are not any more conceivable hubs to visit. One method for playing out the BFS is shading the hubs and navigating as indicated by the shade of the hubs.

There are three conceivable hues for the hub - white (unvisited), gray (went by) and black (wrapped up). Before beginning every one of the hubs are hued in white aside from the source hub that will be shaded gray. To build up this calculation we have to run a similar Mapper and Reducer numerous circumstances (iterative guide reduce).Each emphasis can utilize the past cycle's yield as its information. "Iterative MapReduce and Counters" contains likewise an answer for the most limited way issue (the smallest way between two hubs can be characterized as the way that has the base aggregate weight of the edges along the way. in the event that we don't discuss weighted graph the base weight will be quite recently (the base number of edges). It actualizes this issue using Dijkstra's calculation.

The calculation is intended to make it keep running on a graph that has just a source hub (or as we lean toward call it a "starter hub"), rather our solution needs to has more than a solitary begin hub (i.e. the entirety set of key nodes given in info). At that point we change the graph information's structure with a specific end goal to have the capacity to run the BFS beginning from more than one hub (those will be the key nodes). Likewise we attempted to adjust our information structure as per the lion's share of diagrams. we discovered online which enabled us to utilize our program in bunches of graphs. So we changed the graph contribution from the accompanying configuration:

ID – Neighbors – Distance-From-Source – Color

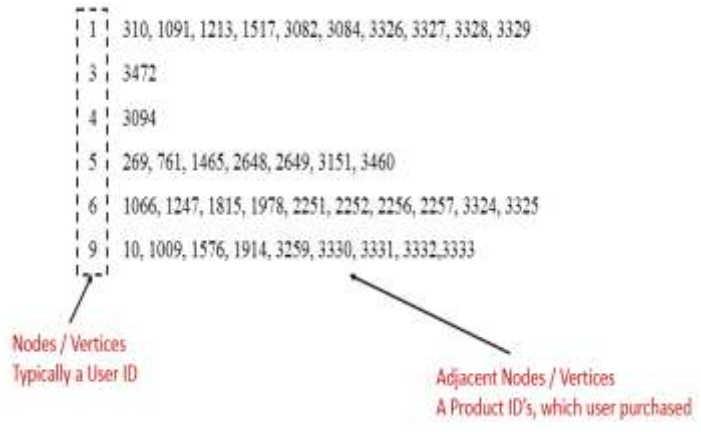

Figure1:-The description of social circle dataset

Figure 2 shows Output to the Reducer,Traversed the miinimum distance .

 ISSN: 2321-8169
445 – 449

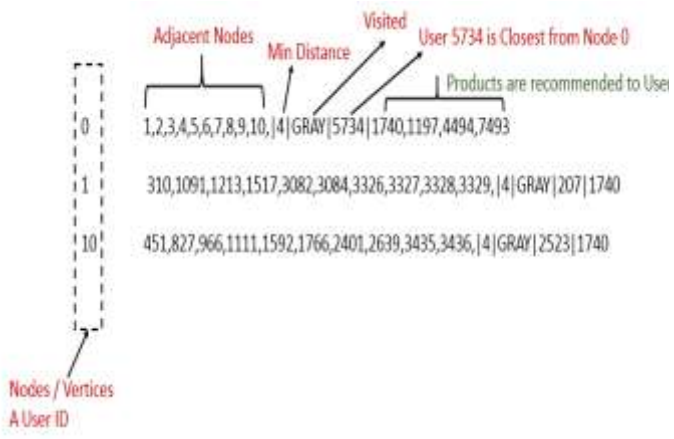

Figure2:-The description of Output Generated by Reduce Function

Various classes used for implementing Recommendation task are given below

A. Base Job

Our project does not need any complex change in the base job of hadoop system, so we choose to keep the Hadoop base employment structure and adjust it with just somewhat alterations without affecting hadoop basic job. BaseJob is utilized to announce the theoretical class JobInfo that contains getter techniques to get the program-particular classes related with Job.

B. Driver Job

This class contains the driver to execute the job and send for the guide/decrease capacities and the fundamental class. The undertaking of the principle class is simply to make the yield catalog and call the run strategy.

C. Nodes

Hub class contains the data about the hub ID, the rundown of nearby hubs, the separation from the source, the shade of the hub (shading could be white for unvisited hub, grY for went to hub and black for the hub which has been gone to by all the key nodes, the parent hub and the starter list (list of the keynodes that went to the hub).

D. Mapper

Mapper Worker is the base mapper class for the projects that utilization parallel expansiveness initially looks calculation. In this class are executed the guide strategy which is in charge of the mapper work.

E. Reducer

Reducer Worker is the base reducer class for the projects that utilization parallel expansiveness initially seek calculation. It joins the data for aindividual hub. The entire rundown of nearby hubs, the base separation from the source, the darkest shading, theparent hub of the hub that is being handled, and starter rundown are resolved in the reducer step.

Figure 1 shows the process of the BFSMpRalgorithm. Firstly, we should use key node and graph as input. Mapper discovers every single Adjacent Node and stamps as Gray then Reducer computes minimum distance and applies BFS calculation from source

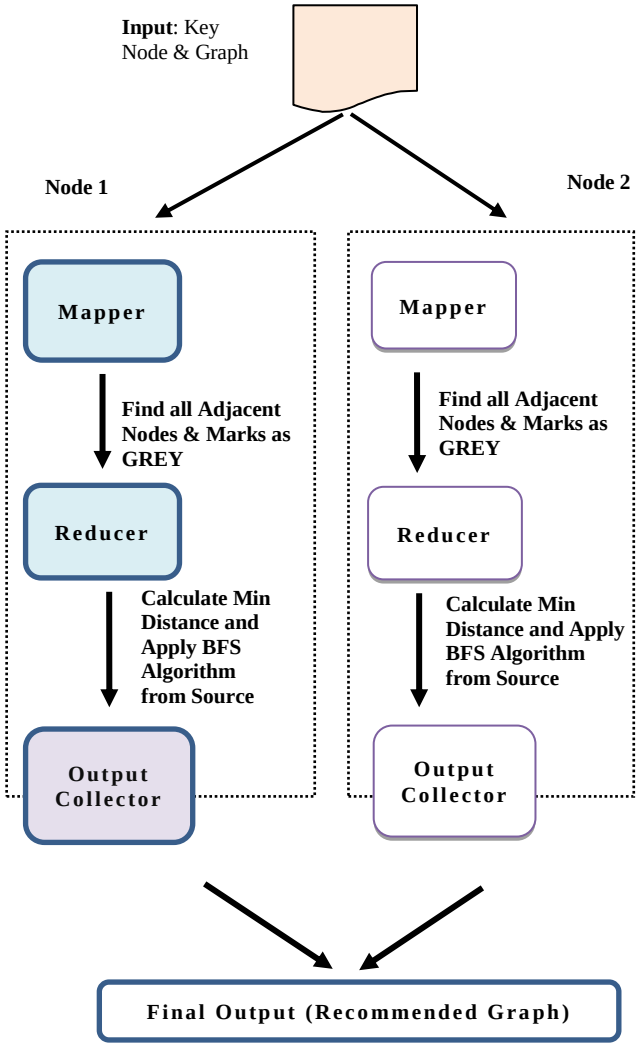

Figure 3:-Work flow of BFS recommndation system

IV. Results And Discussion

After implementation of, the fundamental components of the graph (ID and nearby Nodes) are overseen as a string. This makes conceivable to run the program on various information sorts graphs (i.e. a Node ID might be "1" as can be "Ram" (user name). In any case the biggest piece of graph is spoken to by whole number information sorts with numerical ID. The calculation was prepared using smallgraph (10 to 20 hubs with a most extreme of 50 edges) so we could without much of a stretch check the precision of the outcomes. So we attempted on little string, number and both chart. The program worked appropriately and performing.

BFSMpR took the datasets from the SNAP Datasets Collections [24] that makes accessible an accumulation of

more than 50 huge system datasets from a huge number of hubs and edges to countless hubs and edges. In incorporates informal organizations, web diagrams, street systems, web systems, reference systems, joint effort systems, and correspondence systems.

Figure.4 shows the relationship of different equivalence estimation in based on the time. Here the trial is done on the Comparison among Sequential (non-Hadoop) and Parallel Hadoop Algorithm. the analysis is done on the Hadoop and the non-Hadoop stage. In the Sequential calculation, the program took 40 seconds to finish the execution. Be that as it may, on account of the Parallel(Hadoop stage), it took just a few moments for the execution.

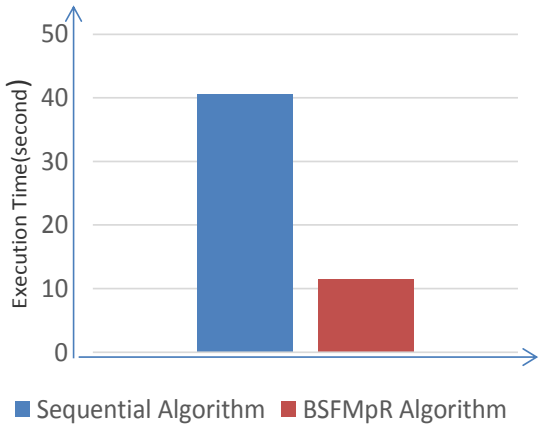

Figure. 4. Comparison between Sequential and Parallel (BFSMpR)Algorithm

Table 1. The performance of job

Test ID	Number of Key-Nodes	Number of Iterations	Time Required
1	30	10	22 sec
2	100	10	32 sec
3	200	10	42 sec

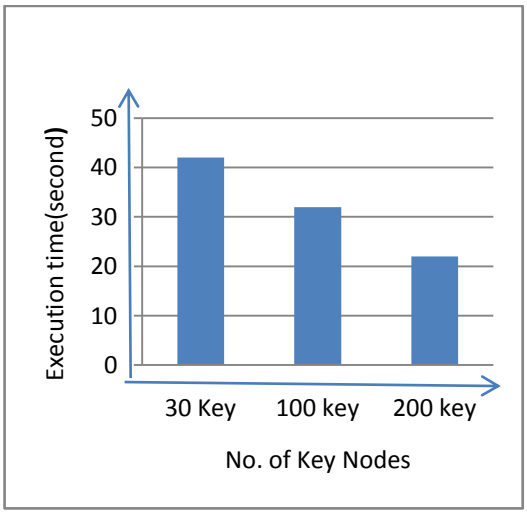

Figure. 5. Execution Time for Key-Nodes

CONCLUSION

This paper introduced BFSMpRalgorithm, which is based on Map Reduce programming paradigm. BFSMpR algorithm outperforms in terms of time with respect to some of the existing sequential mechanisms. There is choice for choosing the key nodes. Hence choosing key nodes plays a very important role to the performance of the system. The performance of the system depends upon the graph and key nodes size. Algorithm recommends product based on the minimum distance of the product with respect to the user. If user A has brought some items X, Y, Z and user C is close with user A and has minimum distance with A, so user C will be likely to recommend product brought by user A.

REFERENCES

[1] J. Dean and S.Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," Commun ACM, vol. 51, no. 1, pp. 107–113, Jan. 2008.

[2] J. H. Hsiao, S. J. Kao, "A usage-aware schedulerfor improving MapReduce performance in heterogeneo us environments," on International Conference on Information Science, Electronics and Electrical Engineering vol 3, pp 1648-1652,2014

[3] "Apache-Hadoop."http://hadoop.apache.org/"

[4] "Applications powered by Hadoop." http://wiki.apache.org/hadoop/PoweredBy.

[5] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The Hadoop Distributed File System," in 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), 2010, pp. 1–10.

[6] M. Newman, "The Structure and Function of Complex Networks," SIAM Rev., vol. 45, no. 2, pp. 167–256, Jan. 2003

[7] Haipeng You, Hui Li, Yunmin Wang, and Qingzhuang Zhao, "An Improved Collaborative Filtering Recommendation AlgorithmCombining Item Clustering and Slope One Scheme", Proceedings of the International MultiConference of Engineers and Computer Scientists Vol 1,pp 313-316, 2015 Honkong.

[8] Dalia Sulieman, Maria Malek and Dominique Laurent, "Graph Searching Algorithms For Semantic-Social Recommendation", IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2012

[9] Yongchang Wang and Ligu Zhu , "Reserch on Collaborative Filtering Recomndantion Algorithem Based On Mahout", Proceedings of The 4th Intl. conf.on Applied computing and Information Technology, 2016.

[10] Dr. Senthil Kumar Thangavel, Neetha Susan Thampi, Johnpaul C I, " PerformanceAnalysis of Various Recommendation Algorithms Using Apache Hadoop and Mahout", International Journal of Scientific & Engineering Research, Volume 4, Issue 12, December-2013

[11] G.R.Bamnote and Agrawal,"Evaluting and Implementing collaborative Filtering Sysystems Using Appache Mahout",International Conference on Computing Communication Control and Automation, 2015.

[12] "Breadth-first search," Wikipedia, the free encyclopedia.

[13] "Breath First Search - Lecture by Rashid Bin Muhammad" http://www.personal.kent.edu/~rmuhamma/Algorithms/MyAlgorithms/GraphAlgor/breadthSearch.html.

[14] K. Mehlhorn and P. Sanders, Algorithms and Data Structures: The Basic Toolbox. Springer Science & Business Media, 2008.

[15] "breadth-first graph search using an iterative mapreduce algorithm.
http://www.johnandcailin.com/blog/cailin/breadth-first-graphsearch-using-iterative-map-reduce-algorithm.

[16] "hadoop tutorial - Iterative MapReduce and Counters." https://hadooptutorial.wikispaces.com/Iterative+MapReduce+and+Counters.

[17] S. Neumann, "Spark vs. Hadoop MapReduce," Xplenty.https://www.xplenty.com/blog/2014/11/apache-spark-vs-hadoopmapreduce

[18] Yonathan Portila,Alexandre Reiffers,Eitan Altman,Rachid EI-Azouzi;"A Study of Youtube Recommendation Graph Based On Meserurments and Stochastic tools" in Proceedings 2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing (UCC),pp 430-435,2015

[19] Brent Smith, Greg Linden;"Two Decades of Recommender Systems at Amazon.com" in IEEE Internet Computing Volume: 21, Issue: 3,pp 12-18,2017.

[20] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," IEEE Internet Computing., vol. 7, no. 1, pp. 76–80, Jan. 2003.

[21] L. Li, D. Wang, T. Li, D. Knox, and B. Padmanabhan, "SCENE: A Scalable Two-stage Personalized News Recommendation System," in Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY, USA,2011, pp. 125–134.

[22] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Itembased Collaborative Filtering Recommendation Algorithms," in Proceedings of the 10th International Conference on World Wide Web, New York, NY, USA, 2001, pp. 285–295.

[23] https://docs.oracle.com/javase/7/docs/api/java/nio/ByteBuffer.html

[24] J. Leskovec and A. Krevl, "SNAP Datasets: Large Network Dataset Collection," Jun. 2015