IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 1, January 2019, ||V-4|| PP 25-28

Hadoop-Based Distributed Dynamic Decomposition for Social

[1]Mr. M. Imayavaramban, [2]Mrs. M. Behima, [3]Mr. M. V. Prabhakaran,
[1]*Dhanalakshmi Srinivasan College of Engineering and Technology*
[2]*Dhanalakshmi Srinivasan College of Engineering and Technology*
[3]*Sree Sastha College of Engineering and Technology*

ABSTRACT
MapReduce may be a programming model and an associated implementation for processing and generating large data sets. The uses dynamic decomposition based distributed algorithm is used which increases the performance of data. The health of the data nodes are verified using health care algorithm. Also the aggregator which is used to group the data from the data node is to be placed correctly by using aggregator placement problem. In this paper, we study to scale back network traffic cost for a MapReduce job by designing a completely unique intermediate data partition scheme. Index Terms: MapReduce, Distributed Algorithm

I. INTRODUCTION

1.1 OVERVIEW

Big Data is a collection of large datasets that cannot be processed using traditional computing techniques. It is not a single technique or a tool, rather it involves many areas of business and technology.

1.2 DOMAIN DESCRIPTION

Big data involves the data produced by different devices and applications. Given below are some of the fields that come under the umbrella of Big Data.

Black Box Data: It is a component of helicopter, airplanes, and jets, etc. It captures voices of the flight crew, recordings of microphones and earphones, and the performance information of the aircraft.

Social Media Data: Social media such as Facebook and Twitter hold information and the views posted by millions of people across the globe.

Stock Exchange Data: The stock exchange data holds information about the ‗buy' and ‗sell' decisions made on a share of different companies made by the customers.

Power Grid Data: The power grid data holds information consumed by a particular node with respect to a base station.

Transport Data: Transport data includes model, capacity, distance and availability of a vehicle.

Search Engine Data: Search engines retrieve lots of data from different databases.

1.3 BENEFITS OF BIG DATA

Using the information kept in the social network like Facebook, the marketing agencies are learning about the response for their campaigns, promotions, and other advertising mediums.

Using the information in the social media like preferences and product perception of their consumers, product companies and retail organizations are planning their production.

1.4 ANALYTICAL BIGDATA

Map Reduce provides a new method of analyzing data that is complementary to the capabilities provided by SQL, and a system based on MapReduce that can be scaled up from single servers to thousands of high and low end machines.

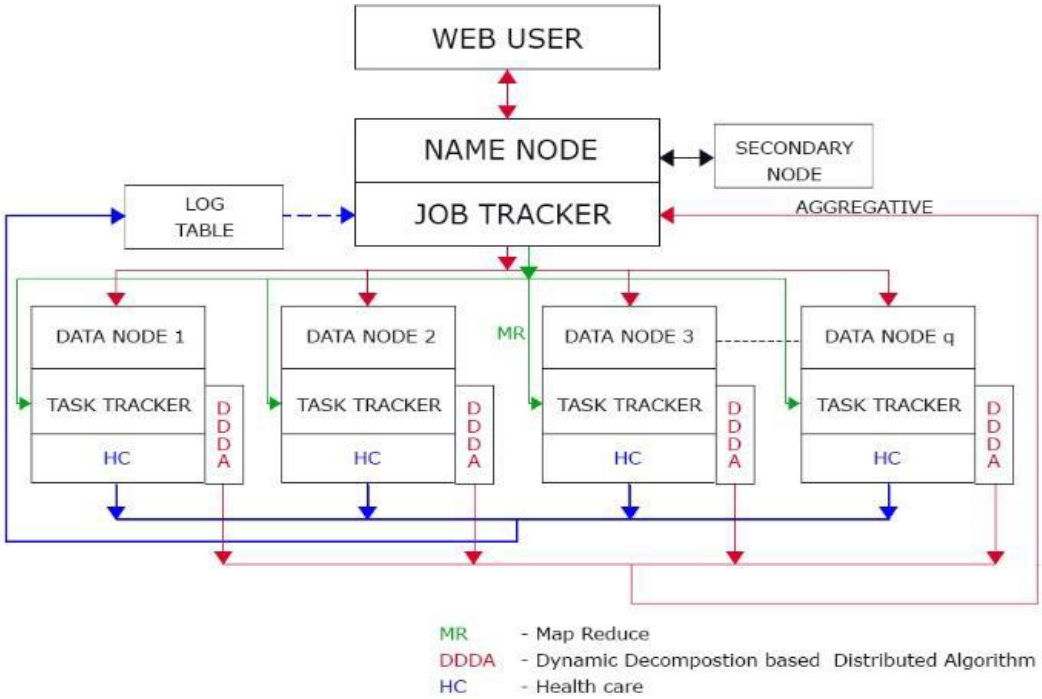
1.5 Hadoop

Using the solution provided by Google, Doug Cutting and his team developed an Open Source Project called HADOOP. Hadoop runs applications using the MapReduce algorithm, where the data is processed in parallel with others. In short, Hadoop is used to develop applications that could perform complete statistical analysis on huge amounts of data.

II. ARCHITECTURE DESIGN

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be the considered to be the most critical in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful

planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

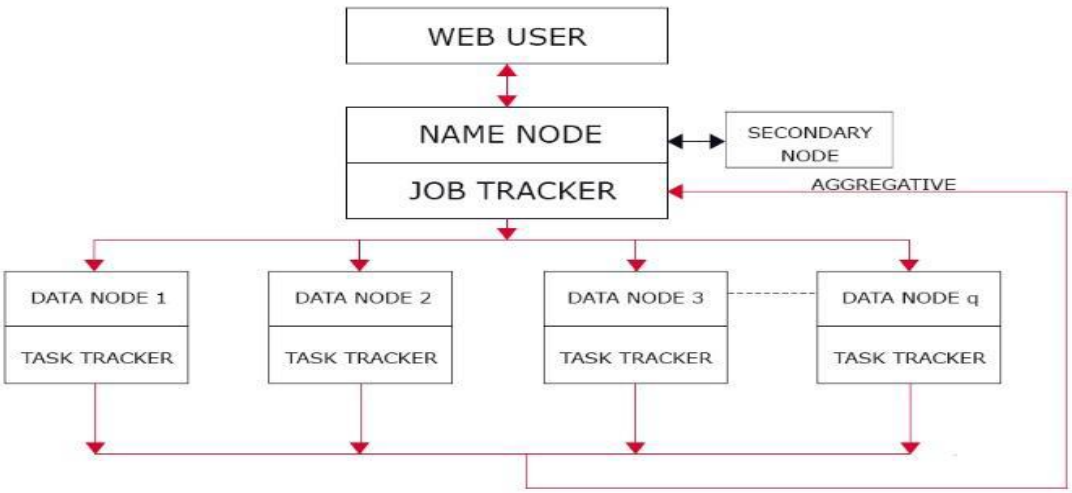


III. MODULES

3.1 Generating Multinode Cluster
3.2 uploading Data To Hdfs Through Policy
3.3 Encryption and Decryption
3.4 Compress Techniques
3.5 DDDA With Framework Monitoring

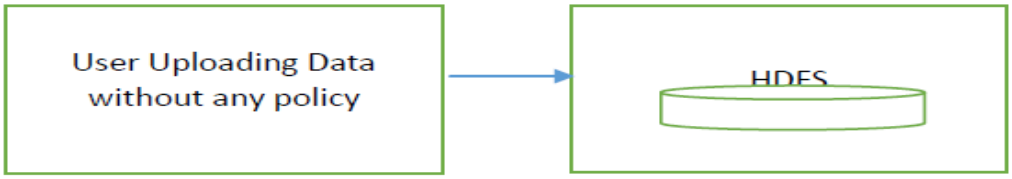
3.1 GENERATING MULTIMODE CLUSTER

In hadoop environment based on the user input no of data node servers will be generated. These data node server have huge amount of storage space than the single data node server in the cloud. These single data node servers are not enough for handling huge amount of data.



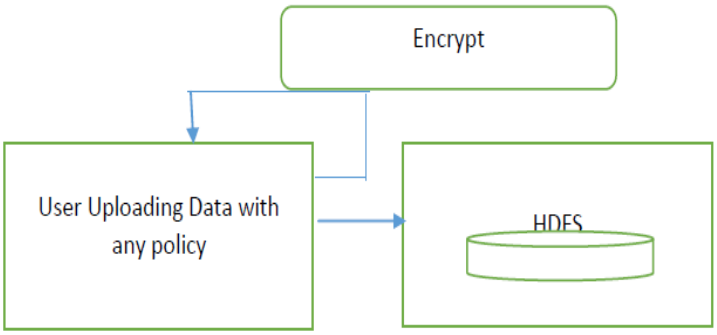
3.2 UPLOADING DATA TO HDFS THROUGH POLICY

In hadoop environment the data are uploaded in 64MB or 128MB block data storage depends on constrain policy are normal, encoding or compression techniques. This is much greater than the normal cloud environment. These data are stored in duplicate copies so that if any data is lost it can be recovered. Before uploading the data are encrypted and compressed for security purposes.



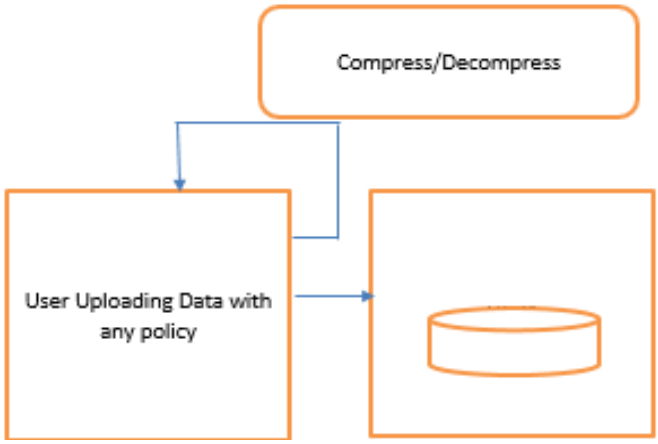
3.3 ENCRYPTION AND DECRYPTION

HDFS Encryption implements transparent, end-to-end encryption of data read from and written to HDFS, without requiring changes to application code. Because the encryption is end-to-end, data can be encrypted and decrypted only by the client. HDFS does not store or have access to unencrypted data or encryption keys



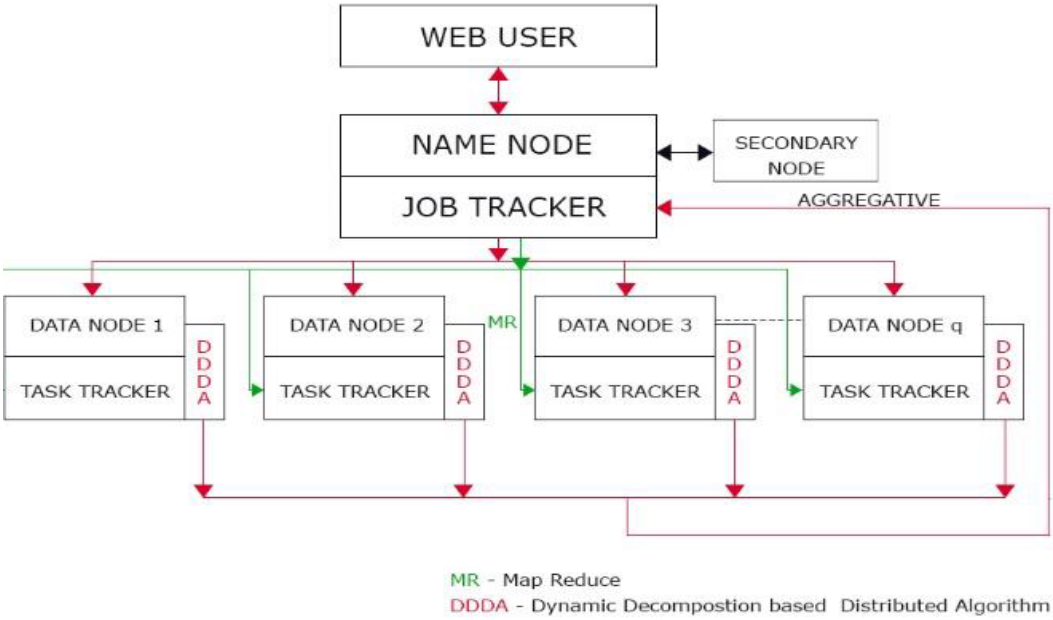
3.4 COMPRESS TECHNIQUES

If the input file is compressed, then the bytes read in from HDFS is reduced, which means less time to read data. This time conservation is beneficial to the performance of job execution.Often we need to store the output as history files. If the amount of output per day is extensive, and we often need to store history results for future use, then these accumulated results will take extensive amount of HDFS space.



3.5 DDDA WITH FRAMEWORK MONITORING

The Dynamic Decomposition based Distributed Algorithm (DDDA) is used to deal with large scale optimization problems for big data application. Also the aggregator which is used to group the data from the data node is to be placed correctly by using aggregator placement problem.



IV. CONCLUSION

CONCLUSION

We explore an integrated network control architecture to program the network at run-time for big data applications. Using Hadoop as an example, we discuss the integrated network traffic control architecture, job scheduling, topology and routing configuration for Hadoop jobs. Our preliminary analysis suggests the great promise of integrated network control for Hadoop with relatively small configuration overhead. Although our discussion has been focused on Hadoop, the integrated control architecture can be applied to any big data applications with a centralized or logically centralized master. Since data aggregation is common in big data applications, the network configuration for aggregation patterns can be generally applied to other applications too

FUTURE ENHANCEMENT

For future work, we are planning to extend the proposed architecture to make it suitable for Big Data analysis for all applications, e.g., sensors and social networking. We believe our study serves as a step towards tight and dynamic interaction between applications and network.

A number of optimizations in our system are therefore targeted at reducing the amount of data sent across the network. redundant execution can be used to reduce the impact of slow machines, and to handle machine failures and data loss.

REFERENCE

[1]. J. Dean and S. Ghemawat,(2008) ―Mapreduce: simplified data processing on large clusters,‖ Communications of the ACM, vol. 51, no. 1, pp. 107–113.

[2]. W. Wang, K. Zhu, L. Ying, J. Tan, and L. Zhang, (2013) ―Map task scheduling in mapreduce with data locality: Throughput and heavy-traffic optimality,‖ in INFOCOM, Proceedings IEEE. IEEE, pp. 1609–1617.

[3]. F. Chen, M. Kodialam, and T. Lakshman,(2012) ―Joint scheduling of processing and shuffle phases in mapreduce systems,‖ in INFOCOM,Proceedings IEEE, pp. 1143–1151**.**

[4]. Y. Wang, W. Wang, C. Ma, and D. Meng,(2013) ―Zput: A speedy data uploading approach for the hadoop distributed file system,‖ in Cluster Computing (CLUSTER), IEEE International Conference on. IEEE, pp. 1–5

[5]. L. Fan, B. Gao, X. Sun, F. Zhang, and Z. Liu,(2014) ―Improving the load balance of mapreduce operations based on the key distribution of pairs,‖ arXiv preprint arXiv:1401.0355,.