Scientific Journal of Impact Factor (SJIF): 4.14
e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 3, Issue 3, March -2016

INFORMATION RETRIEVAL FROM BIG DATA FOR SENSOR DATA COLLECTION

Varsha Jawale , Swapna Aware, Smita Kolase , Mr.Amit Sonawane
Department of Information Technology, MIT academy of engineering
Department of Information Technology, MIT academy of engineering
Department of Information Technology, MIT academy of engineering
Asst. Professors, Department of Information Technology, MIT academy of engineering

Abstract *— Hadoop is platform of big data it is a collection of large datasets that cannot be processed using traditional computing techniques. Hadoop partition the work to bunches of machine and organizes the work among them. It has two segments Map and Reduce. HDFS give high accessibility of information to client application and it is intended to store huge measure of information dependably. Map Reduce is additionally a product structure for the investigating and handling huge information set into required sought yield. This paper essentially concentrate on how the copies are going to oversaw in the HDFS for giving high accessibility of information under tremendous measure of computational necessity. This paper additionally concentrate on straightforward information model gave by sensor information which gives customers dynamic data of area over information format and we depict the outline a group and execution of investigation of sensor information utilizing cloud.*

Keywords- Hadoop, HDFS, Map Reduce, Sensor Devices, Cloud.

I. INTRODUCTION

The Hadoop is parallel access to information that can live on a one hub or a large number of hubs. HDFS makes number of numerous reproductions of information pieces and disseminates them on figure hubs all through a bunch to empower with suitable dependability and high fast calculations. Hadoop is uses a scale out piece design that makes utilization of merchandise servers arranged as a group, where every server has reasonable inside plate drives. Hadoop applications utilized HDFS as the essential stockpiling framework and Hadoop have own information structure. Information in Hadoop is separated into different squares and spread all through a group. Once that happens after that Map Reduce errands can be completed on the littler shared of information that might make up a vast dataset in general, in this way achieving versatility required for huge information handling. It can complete some sort of programmed repetition and come up short over make it famous for the new up and coming organizations which are searching for information stockroom bunch systematic arrangements. Map-Reduce are a programming model furthermore related execution for handling and producing vast information sets. Clients characterize the guide work that procedures a key-esteem pair to create an arrangement of halfway key-esteem matches and decrease work that unions every single middle quality connected with the same moderate key.

II. LITERATURE REVIEW

1. Distributed Mining Algorithm Using Hadoop On Large Data Set
The main objective of data mining is to discover knowledge from large databases. The discovered knowledge helps in decision making and The Apriori like method suffers from two main problems. One is main memory has to be large to hold all candidate item sets. Second it scans the database multiple times.

2. MapReduce: Simplified Data Processing on Large Clusters
We have learned several things from this work. First restricting the programming model makes it easy to parallelize and distribute computations and to make such computations fault-tolerant. Second network bandwidth is a scarce resource. A number of optimizations in our system are therefore targeted at reducing the amount of data sent across the network: the locality optimization allows us to read data from local disks, and writing a single copy of the intermediate data to local disk saves network bandwidth. Third redundant execution can be used to reduce the impact of slow machines and to handle machine failures and data loss.

3. A Survey on Big Data in Real Time

We have entered an era of Big Data. Through better analysis of the large volumes of data that are becoming available, there is the potential for making faster advances in many scientific disciplines and improving the profitability and success of many enterprises. We have only begun to see its potential to collect, organize and process data in all walks of life.

4. Power Grid Time Series Data Analysis with Pig on a Hadoop
Cluster compared to Multi Core Systems We compared multi-core with distributed data processing for simple statistics analyses. For this purpose we developed the statistical analysis functionality in Java for both the Hadoop cluster using the Pig-API and the local environment.

III. SURVEY OF PROPOSED SYSTEM

This project basically for the generate a big data to apply a algorithm to fire a query to get knowledge discovery. It will be for decision making. Single node which is used to collect sensor data. We try to map geographical coordinate from sensor device. These co-ordinates are change dynamically as the sensor device is moving geographically. The main intention for collect coordinate is to keep track that device. A cloud based sensor data processing system can process data in two ways- First sensor network sense data and forward it to the cloud for processing and secondly cloud data processing system first process query, subdivide the query and forwards into sensor network. Cloud is used because it can access anywhere anytime and it is best suited for big data analyze. As we are using cloud it is beneficial data collect dynamically. In response of query sensor nodes sends only required data. This required data is to analyze for a query which help for knowledge retrieval .

IV. Mathematical Model

Let S is the Whole System Consist of

S= {I, P, O}

I = Input.

I = {U, Q, D}

U = User

U = {u1,u2….un}

Q = Query Entered by user

Q = {q1, q2, q3…qn}

D = Dataset.

P = Process:

Step 1: Collect sensor data.

Step 2: Apply association rule on transaction database.

Step 3: Simply search using Map reduce hadoop algorithm.

Step 4: Generate big data.

Output: The output will be the response of the user query

V. SYSTEM ARCHITECTURE

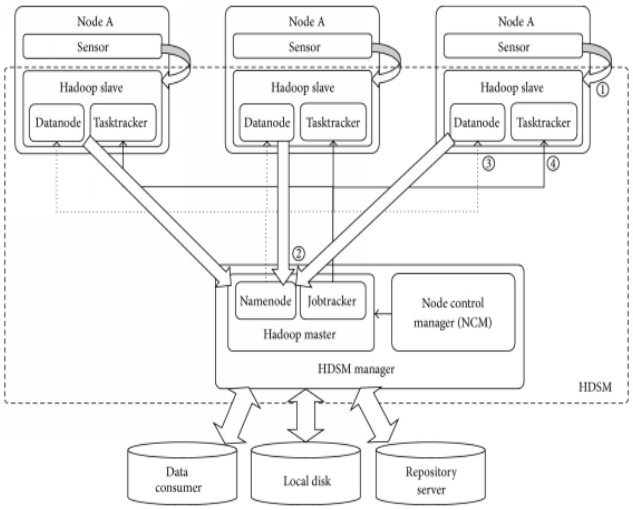

Fig 1. System architecture

VI. CONCLUSION AND FUTURE WORK

Proposed framework is pointed towards information gathering in element environment. Progressively to track the area of adaptable from geographic co-ordinate and store it on slave of Hadoop bunch utilizing map-decrease engineering learning is recovered which will supportive in crisis case. Further utilizing this design information is recovered which will supportive to the client instance of crisis viz. Mischance

VII. REFERENCES

[1] Dean, J., Ghemawat, S.: MapReduce: a exible data processing tool. Communications of the ACM 53(1): 72-77 (2010).
[2] M. Yuriyama and T. Kushida, "Sensor-Cloud Infrastructure – Physical Sensor Management with Virtualized Sensors on Cloud Computing," sept. 2010, pp. 1-8.
[3] Stacey Higginbotham. (2010, September) Sensor Networks Top Social Networks for Big Data. [Online]. http://gigaom.com/cloud/sensor- networks-top-social-networks-for-big-data-2/ [Accessed on: 2012-06-08]
[4] Divyakant Agrawal, Sudipto Das, and Amr El Abbadi, "Big data and cloud computing: current state and future opportu- nities," , New York, NY, USA, 2011, pp. 530-533. [Online]. http://doi.acm.org/10.1145/1951365.1951432
[5] Raquel, M.M., Rose, B., Butt, A.R. and Nikolopoulos, D.S.(2009) Supporting map reduce on large-scale asymmetric multi-core clusters
[6] Talia, D. and Trun_o, P. (2010) How distributed data mining tasks can thrive as knowledge services, Communications of the ACM, Vol. 53, No. 7, pp.132137.
[7] A. Bifet and E. Frank. Sentiment knowledge discovery in Twitter streaming data. In Proc13th International Conference on Discovery Science, Canberra, Australia, pages 115. Springer, 2010.
[8] B. Pang and L. Lee. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2)

AUTHORS

Varsha Jawale, Pursuing BE. *Department of Information Technology, MIT academy of engineering.*
Swapna Aware, Pursuing BE. *Department of Information Technology, MIT academy of engineering*
Smita Kolase , Pursuing BE. *Department of Information Technology, MIT academy of engineering*
Mr.Amit Sonawane , *Asst. Professors, Department of Information Technology, MIT academy of engineering* - *amit29sonawane@gmail.com*

