

Trident Technology Usage in Query Search time by using Cache Memory

K.Mamatha, J.Saritha

Abstract— *Big data is technology which has the users increasing day-by-day. This technology is now using in all the fields for storing the large dataset. Large datasets cannot be handled by normal database. So we are using this big data. Here in this paper we are using trident technology to avoid the issues searching query in hadoop like time. So by using cache memory is used with trident technology to perform the search operation faster than the hadoop. Here we are using the SHA algorithm while clustering the dataset.*

Key Terms: *Big data, Hadoop, Clusters, Nodes, Distributed server.*

I. INTRODUCTION

Big data can hold both the structured and unstructured data. It is used to store the large amount of the data. Now in query organisation is using this big data to store their large amount of information about company and also about the Hadoop it is using in every field. It was

present in big data. Apache hadoop work with different languages like ruby, pig, hive and some other. It works on the virtual machine. In this hadoop we have a mapper and reducer. Mapper and reducer will work by combining each other work. With the help of these two hadoop will run and solve all the queries. Existing techniques like Hadoop and spark but it can also process big data but they lack of query search operations and Hadoop it will take much time to partition data between map and reduce face and Hadoop does not create feature vector to perform query operation and has no support to create cluster to store similar data on one node. Feature vector is nothing but information describing objects. In this paper mainly proposed on one group of class applications: those that recycle a working group of data across various parallel operations [1]

Trident is nothing but distributed analytical platform. The modelling operation by assume data with high quality searching operation applicable and also advice in the making successful model. Cassandra it is also one type of distributed storage system. It aims to run on top of a substructure of hundreds of nodes. It manages overcome of this paper reliability and scalability of the software systems based on this service [2]. projects. Big data mainly here hadoop in it. Hadoop has mapper and reducer. Mapper is used for assigning the jobs or tasks and reducer process the job and

with the both combination of these results we will get the final result done.

II. RELATED WORK

2.1 Big data:

Big data has 4 V's present in it. Volume explains nothing but the size of the data which is stored. Variety explains the nature of data and also the type of the data which was stored. Velocity explains the speed of the data generation and meeting the challenges. Veracity explains about value and quality of data that is stored. North American Disease Spread Model it was developed with broad international carries to help policy evolution and disease incursions [4].

Map Reduce and its forms have been highly victorious large-scale data in-depth applications item clusters. In this discuss about recover a set of data across involving several operations. Those comprise many repetitious machine learning algorithms as well as synergistic data analysis .condition high-performance [6]. LIBSVM it is a library for support Vector Machines. It has obtained large popularity in machine learning and other areas [8]. It has been replaced by high-level breeding. Many local have been frightened by disappearance and the loss of their native genetic diversity [5].

Trident it can mainly using overcome feature vectors Hadoop and Spark it will not create feature vectors and does not create cluster similar data on one node. Hadoop it will take more time and lack of query operations. These drawbacks overcome in trident. Machine learning it is a form of AI that authorizes a system to learn from data rather than through explicit programming. Geo hash it is nothing but a public domain geocoding system. It is divided in to a short string of letters and digits. It can be performed hierarchical data structure. It is divided in to buckets of framework shape [3]

III. PROPOSED WORK

In this proposed work we are going to explain about trident. Trident it is mainly focussed on storage level and efficient query search operations on big data. Distributed file systems does not support query results fast. Training data also performed. Big data it will not create feature vector and it will take more time efficient. Python Data Analysis Library it is an open source, BSD- permit library high-performance [6].

Revised Version Manuscript Received on 16 September, 2019.

K.Mamatha, M. Tech Student, Dept Of Computer Science And Engineering Jntua College Of Engineering Pulivendula, Andhra Pradesh, India (Email: kmamatha839@gmail.com)

Dr .J.Saritha, Assistant Professor, Dept of Computer Science And Engineering Jntua College Of Engineering Pulivendula, Andhra Pradesh, India

3.1 Advantages:

- A fast query-driven approach is support recapture of training data.
- Modelling advice provided by automated dimensionality trimming.

3.2 System Architecture:

In this concept to store and efficient query search operation on big data. Hadoop and Spark it can also process big data but they lack of query search operations on big data, and it will take large time to partition data between map and reduce face and hadoop it cannot be created feature vector by reducing their dimensionality. It has not create cluster similar data on one node. To overcome above issues trident has support this following features

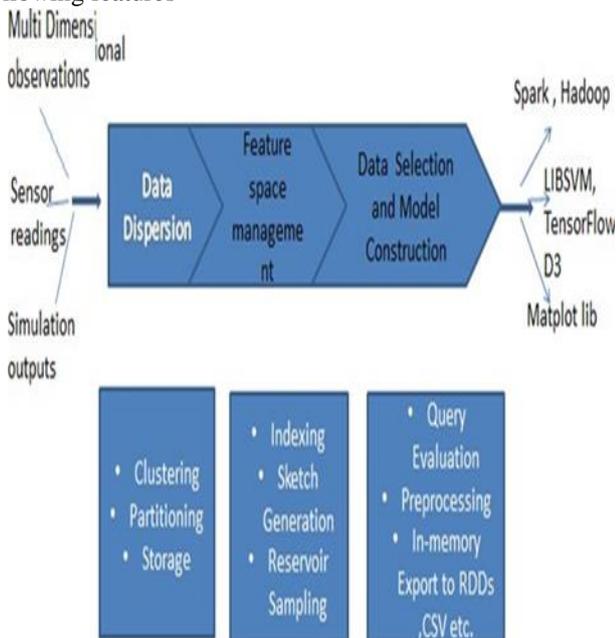


Figure 1 Trident architecture

Above figure 1 explains about the trident architecture which helps us to perform the present work. Architecture contains:

3.2.1 Data dispersion:

Here data will partition on different nodes by forming clusters, each node will have similar data only and node will store data in memory. It can be divided in to storing, clustering and partitioning.

3.2.2 Feature Space Management:

To support query operation trident will create index tree with all features available in records or dataset.

3.2.3 Data Selection and Model Construction:

With entire training data and index tree a model will be constructed and whenever user issue any query then trident will perform search operation on model and give result back to user. In this datasets can author has introduced three datasets. Animal disease spread model it is not available on the internet. North American disease spread model it is also not available on the internet. Here in this paper we are taking the atmospheric dataset for the query search.

3.3 Secure Hash Algorithm:

It is a cryptographic hash function. It is used in the initiate work to show the details of the clustered data in the nodes with names of the file. Top level algorithm is known as SHA-3 algorithm.

IV. RESULTS AND ANALYSIS

Storing and Efficient query are the important tasks. In order to Storage data and similar cluster values store on data node we are using trident. In this result and analysis we perform the operations of taking dataset, uploading dataset, forming clusters, searching query and performance analysis between normal search time and cache search time.

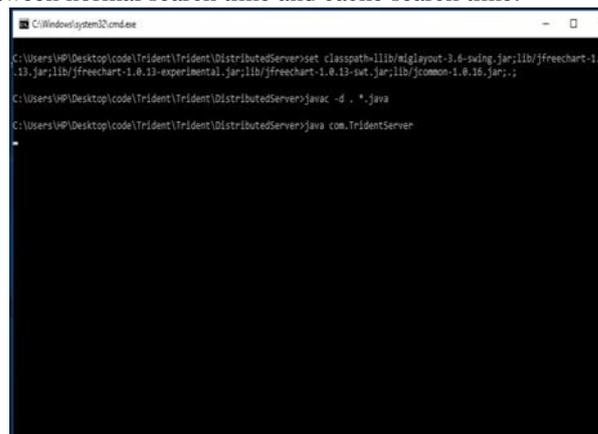


Figure 2 Distributed Server running

Figure 2 shows the distributed server run 'run.bat' file from Distributed Server to get screen and let it run. This application is responsible to read data set value and then create index to generate cluster and perform query operation and then all similar cluster data sent to each node. Here we run the run.bat for distributed server.

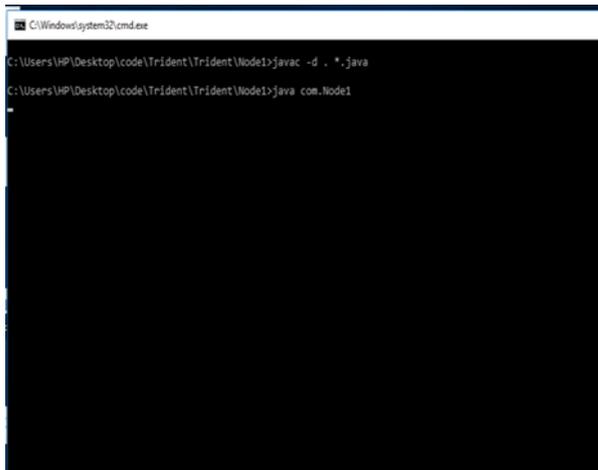


Figure 3 Node1 Server running

Figure2 shows the running of 'run.bat' file on Node1 Server. This is a Storage application which accepts similar partition cluster data from Distributed Server data store in memory. Here we run the run.bat for node 1.



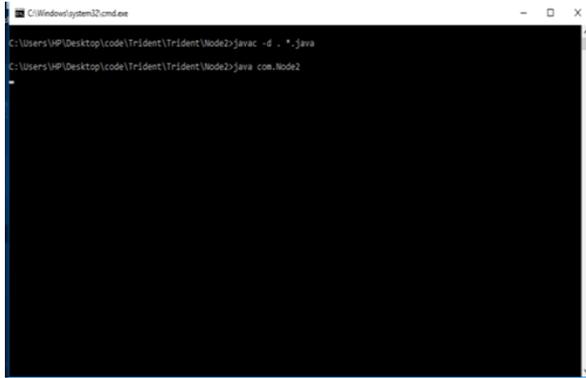


Figure 4 Node2 Server running

In this above figure3 shows this is another storage node which accepts data from Distributed Server and store in memory. Here we run the run.bat for node 2.

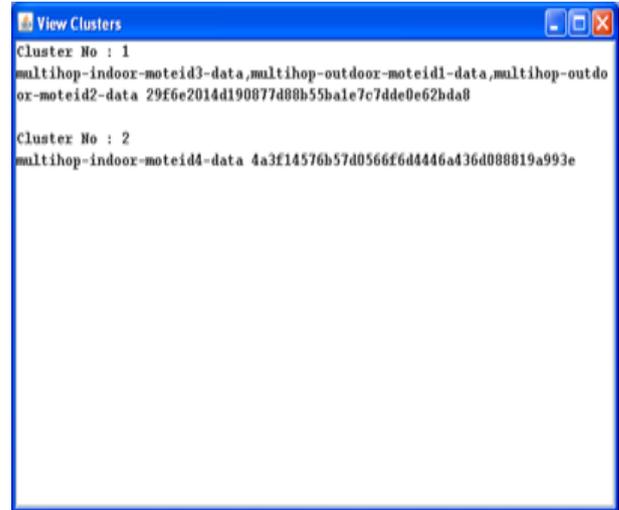


Figure 6 clusters divided

In above figure 6 it shows clusters that are separated and stored in the node1, node2. Clusters formed has three similar files which are stored in two nodes. One cluster has similar data and other cluster contain non-similar data. Here SHA is used for displaying the files along with their names.

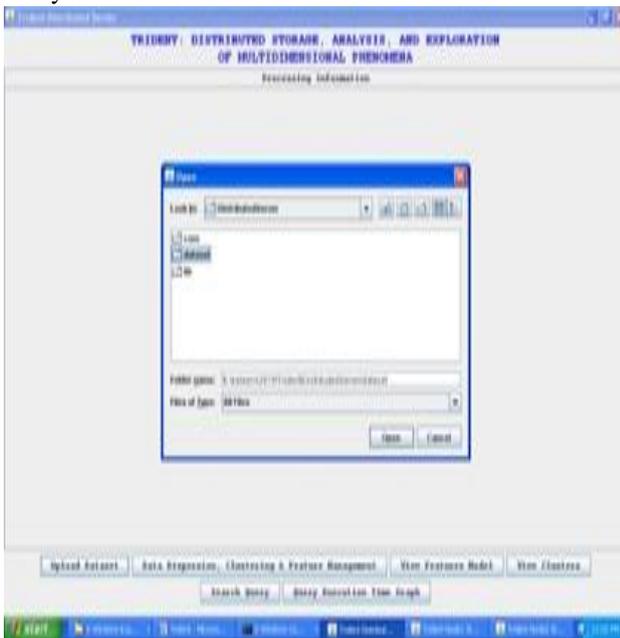


Figure 5 Uploading Dataset

In figure 5 diagram shows uploading of the data set. After uploading the dataset click on Data Dispersion, Clustering and Feature Management button to generate clusters and feature vectors. The total no of dataset files available and total no of clusters generated and the name of similar file in each cluster is also displaying with SHA code and we can see which file sent to which node for storage. Now click on 'View Feature Model' to view features value generated from dataset. Features will be > 0 if two files contains similar data otherwise the value will be 0 and by seeing this value we can see weather two files contain similar data or not. In above table each row represents features values from each file.



Figure 7 Searching query

In above figure 7 we are going to search the query from the dataset which was clustered and present in the nodes. After the query was searched we can we view time variation of searching in normal search time and cache search time.

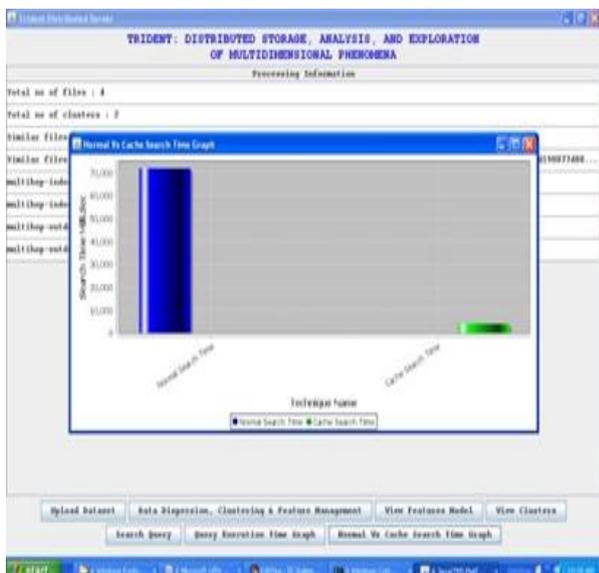


Figure 8 Normal search time vs. Cache search time

In above figure 8 it shows the time variation between the normal search time and the cache search time of the query. By this we can say that the cache search time takes less time when compared with the normal search time.

V. CONCLUSION AND FUTURE WORK

The proposed concludes that it better to use the cache search time rather than the normal search time as it takes less time for the searching query. The query once searched is stored in the cache memory it recall immediately when the same query is given but it was not the same process in normal search time the query is not stored in the cache. As for the future work we can include that trident for supporting the visualizations.

REFERENCES

1. M. J. Franklin, S. Shenker, M. Jaharia, I. Stoica and M.Chowdhury,” Spark Cluster Computing with working sets”.
2. P. Malik and A. Lakshman “Cassandra”
3. G. Niemeyer Geo hash published on 2008.
4. C-J. Lin and C-C-Chang “Libvsm”.
5. <https://www.r-project.org>.
6. <http://pandas.pydata.org>.
7. M. Schoenbaum et al., N. Harvey, A. Reeves “NAAM Model”.
8. V. Michael, F. Pedrogosa, B. Thirion, G. Varoquaux, M. Blondel, V. Dubourg, M. Brucher, O. Grise