ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

A Significant Big Data Interpretation Using Map Reduce Algorithm

Shruthi Bariki[1], Pushpa S Tembad[2]

M.Tech Student, Dept. of Computer Science, STJIT, Ranebennur, India[1]

Associate Professor, Dept. of Computer Science, STJIT, Ranebennur, India[2]

ABSTRACT: Ontology involved in growing data based on semantic web has brought mind-blowing performance in efficiency and scalability. Large ontologies cannot be processed with centralized reasoning methods; hence we implement distributed reasoning methods which improve the scalability and performance. This implementation is based on distributed and incremental reasoning methods for large scale ontologies where we make use of Map-reduce which involves high performance reasoning and runtime searching. Reasoning process can be simplified by making use of EAT and TIF along with it storage is also largely reduced. The main thing here is we make use of system that is implemented on the Hadoop framework where in results obtained from experiments show the effectiveness of the proposed work.

KEYWORDS: Big Data, MapReduce, Ontology reasoning, Semantic Web, RDF.

I. INTRODUCTION

In today's era, Semantic web data is increasing rapidly. Various applications have begun which is due to fast data growth in semantic web. Like, Health-care, business process management, expert structures market place, web facilities arrangement and cloud system management. As we know that web data is increasing day by day, Likewise, semantic web data has also increased from million to billion triple. According to historical analysis its growth hasn't stopped yet. Here we come across one major issue that is searching knowledge in this big data set. Ontology: Ontology is web grid, which encompasses vocabularies, meaning and definition of some domains. Since, it provides a machine operated and formal model of the domain. Gathering of information based on knowledge is done using RDF's which is represented ontologically. Object, subject and predicate is the triplet form of representation. Here Resource, Predicate denotes parts of resource and it also denotes the relationship between resource and object. Anyways, existing distributed reasoning method points on developing RDF closure's. Preparing this RDF closure itself involves much time and consumes lot of space. Hence this is a drawback. In order to overcome this we use a technique called large scale RDF based on distributed and incremental data set through Map Reduce.

In this particular project we use Map-Reduce technique for inference method. Here we come across two functions they are: Map () and Reduce () functions. Map () performs the sorting and filtering whereas Reduce () helps in summarizing the operation. The "Map-Reduce System "Provides helps in organizing the servers, which runs multiple tasks in parallel and performs data transfer among various system which provides fault tolerance and redundancy. In order to process large data set it is more convenient to use the Map-Reduce programming model. Good Map-Reduce algorithm helps in optimizing the communication cost. For storing the RDF triples in efficient way, we use two concepts they are: IF (Transfer Inference Forest) and EAT (Effective Assertional Triples).These Two will strongly help in reducing the storage and it will also simplify the reasoning process By using TIF and EAT, one can avoid computing RDF closure with which user queries can be quickly answered. So, processing time will be reduced, which will be ranked as best method till today compared to the existing method. Here we make use of Hadoop platform to implement our prototype. Hadoop is open source software where billion triple challenges are being performed with different methods which are considered as a benchmark data. It can tolerate high level fault tolerance and is also capable of detecting and managing failures. It can successfully run and monitor Map-Reduce application on cluster. In order to distribute and implement scalability in storing the data we make use of HBase(Hadoop Base).It provides the read write access to very large table which encompasses billion of rows and columns on cluster.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

II. RELATED WORK

[1]Computing RDF closure is one of main criteria in distributed reasoning method, but it takes too much of time and space. It is important to build relationship between fresh RDF triples and older ones but it does not attain success in reproducing the relations between them.
In order to improve the performance of context we implement two fuzzy implication engines were proposed which was based on knowledge representation model. Concept parting policy and semantic implication engine made use of an algorithm called multi-phase forward-chaining algorithm which helped a lot to solve the semantic inference difficulties for instance in e-market events. Map-Reduce are based on distributed reasoning method which is useful to compute RDF graph. As the procedure is being carried out on Hadoop the disadvantage of Map-reduce is being stressed hence came the Map resolve method to solve more expressive reasons .When ontology is being updated and the volume of data increases, solving this kind of problem it is very much necessary to re-calculate the entire RDF, whenever new data arrives. Reasoning methods has to be improved, in order to save time. RDF closure and RDF datasets were calculated using scalable implication method. They too modified their status as incremental perspective to process the declaration rendering. where in this we face some problems because of drawbacks of this existing method they are as Current reasoners did not reuse old results obtained which will lead to increase the process time, Old results were not being used by current reasoned which led to increase the time to process, Some ontology principles cannot follow assuming axioms, Classifying the whole ontology is cheaper than classifying large fragment, Open knowledge bases and assumptions are not real, In pattern recognition tasks were not allowed by tree language interference algorithm.

Thus in order overcome these drawbacks of the existing method we propose an one method in that proposed method [2]From the original Resource Description Framework data the ontology information has gathered. In order to decrease the size of the contributed data the triples indexing module and dictionary encoding module are encodes the all triplicates into an exclusive and small identifier. Make use of k means clustering algorithm we run a Map-Reduce algorithm which compress a huge amount of RDF data in side by side. Optimizing the classified data on database engine a simple pattern can be prepared with it. To fetching the interior set of data finally Query retrieval and Construction is allocated. Queries Delivery is based on map reducing function performance. To gathering false positive and false negative performance evaluation we are constructing recall and precision. From this proposed method we can acquire several advantages these are; now we can simplify the reasoning process and generally reduce the storage, We can solve real world application challenges, By making use of previous versions of ontology we can reuse the information, Description of two things is given by RDF data is tripl.

III. PROPOSED ALGORITHM

A. *Design Considerations:*

In this particular project we use Map-Reduce algorithm for inference method. Here we come across two functions they are: Map () and Reduce () functions. Map () performs the sorting and filtering whereas Reduce () helps in summarizing the operation. The "Map-Reduce System “Provides helps in organizing the servers, which runs multiple tasks in parallel and performs data transfer among various system which provides fault tolerance and redundancy. In order to process large data set it is more convenient to use the Map-Reduce programming model. Good Map-Reduce algorithm helps in optimizing the communication cost.

B. *Description of the Proposed Algorithm:*

Aim of the proposed method is to reduce the storage. Our proposed approaches validation could be implemented on the platform of `Hadoop`. This is mostly used for enabling the Map-reduce Technology. Figure 1 shows the prototype system architecture. [2]The IDIM modules are the core of the system, this receives the input of incremental RDF datasheets, then makes triple processing and interacts with HBase to read and store the intermediate results, then forwards the query output for the end users.

To store the encoded ID, CTIF, PTIF, DRTF, CEAT, and PEAT we have implemented six numbers of HBase tables. The framework of Hadoop involves Map-reduce implementation which allows large datasets distributed processing across computers clusters through the models of simple programming. This could extend from one server to thousands of machines. Every server proposes the storage and computation, and manages details of execution such as job scheduling, data transfer and management of error.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

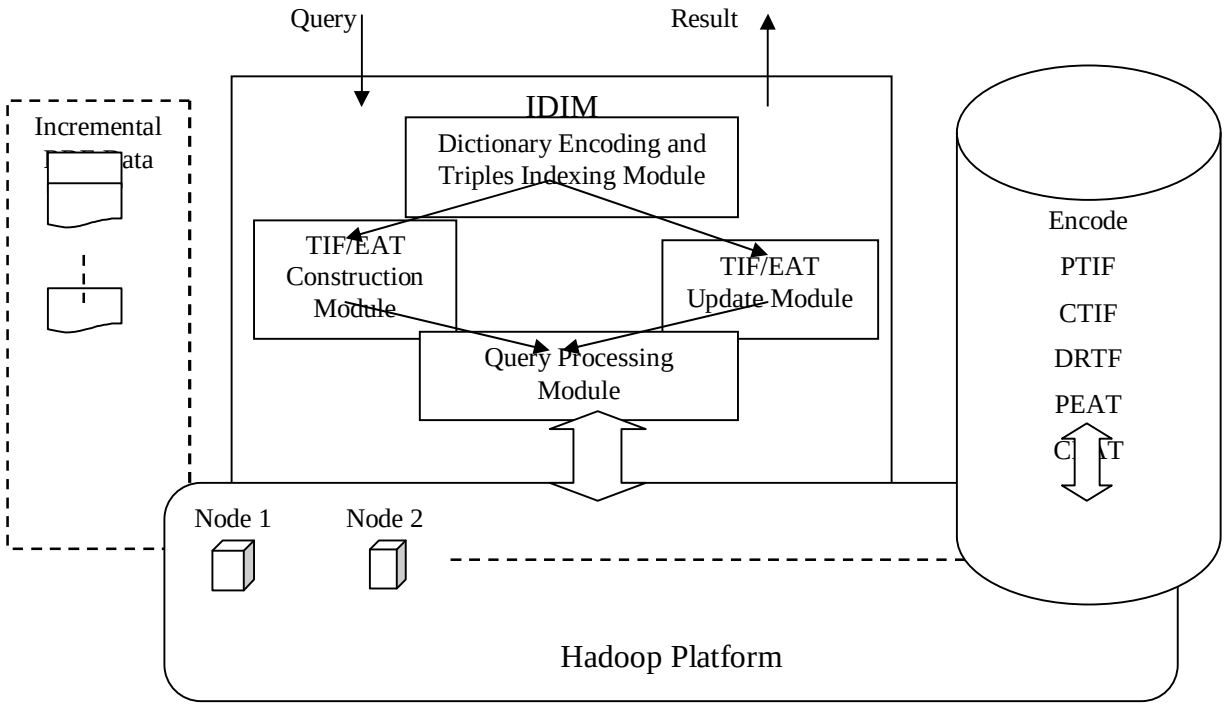

Fig 1: Proposed System Architecture

IV. PSEUDO CODE

A. // Pseudo code...//

Step 1: Start
Step2: IDIM Modules which is the main core of our proposed system takes input as incremental RDF data.
Step 3: After receiving incremental data sets as inputs it makes triples.
Step 4: Then next IDIM modules starts processing of these triples and performs the reasoning by means of a set of MapReduce programs.
Step 5: During this process it interacts with the HBase to read and store the intermediate results, then forwards the query output for the end users.
Step 6: Stop

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

B. Flow Diagram of Pseudo code

Figure 2 shows the flow diagram of the pseudo code.

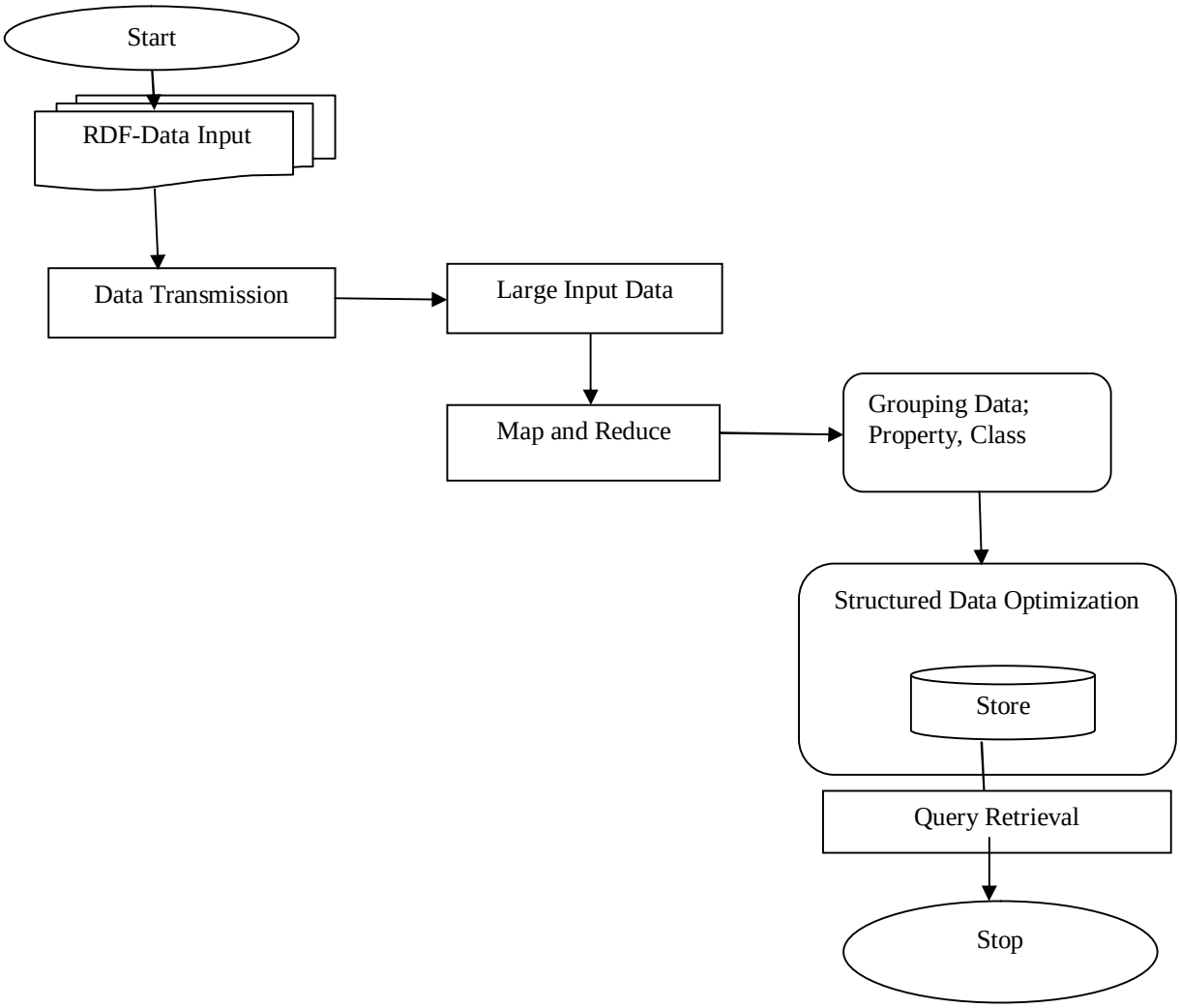

Fig 2: Flow Diagram

As shown in the above figure 2 RDF data is fed as input for data transformation, once data transformation over we get an large input data where this as fed as input to map and reduce algorithm. Map and Reduce algorithm group the data based on property and class, after this it performs structured data optimization, and finally respective query output will be delivered to end user.

V. EXPERIMENTAL RESULTS

Once we fed large input data to map and reduce algorithm, this algorithm performs grouping and optimization, as a result of these tasks of map and reduce algorithm we get a mapped result and reduction output as shown in the figure 3 and figure 4.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

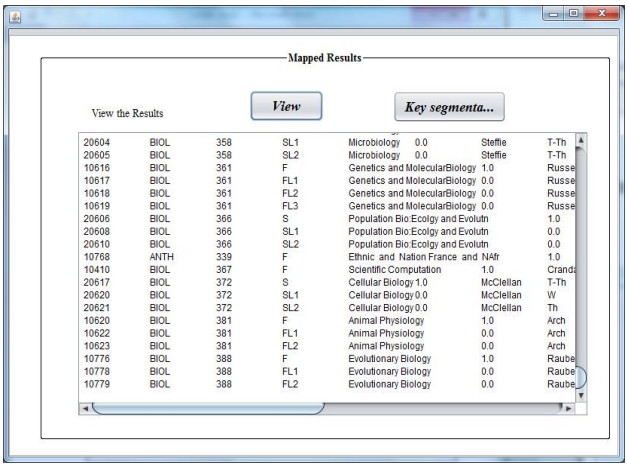

Fig 3: Mapped Result

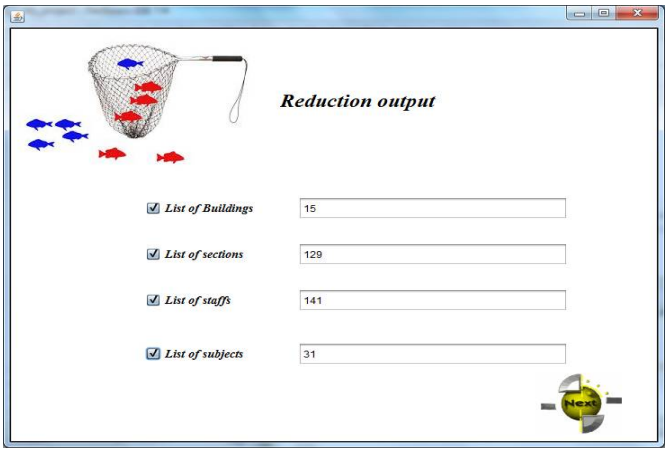

Fig 4: Reduction Output

Following figure 5 shows an performance of the proposed system mainly the performance of the map and reduce technique, in that red colour shows an map reduce recalculation and blue colour shows an map reduce precision and figure 6 shows time graph where in that red colour shows how much time the proposed system takes for producing query response and the blue colour shows time requirement of the existing system for producing query response. From this graph we say that the proposed system takes lesser time for producing query response as compared to time requirement of existing system for producing query response.

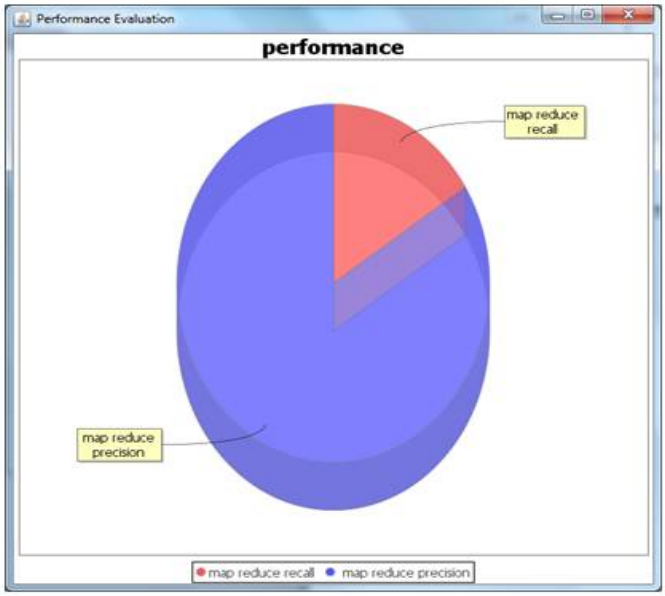

Fig 5: Performance page

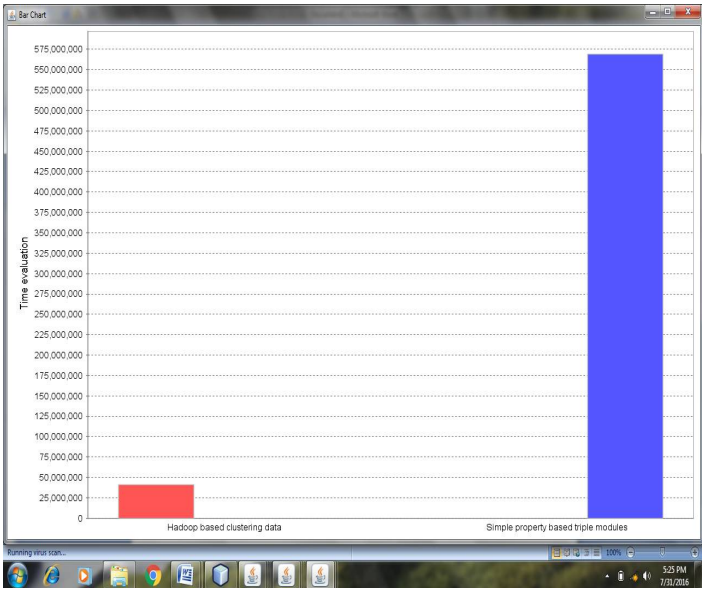

Fig 6: Time Graph

VI. CONCLUSION AND FUTURE WORK

As Big data is the major concept of today's era and here the data involved is of huge volume and the task complexity, the reasoning on web scale has become progressively challenging. On each update the full reasoning on entire dataset is much time taking to be practical. Map-reduce and Hadoop uses minimum nodes like for instance eight

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

nodes. BTC benchmark is the one where the system is being checked and it covers all the aspects. In this case we will test more datasets and make use of more number of nodes where there is possibility to extend IDIM to OWL which involves other ontological vernaculars.

REFERENCES

1. M. S. Marshall et al., "Emerging practices for mapping and linking life sciences data using RDF—A case series," J. Web Semantics, vol. 14, pp. 2–13, Jul. 2012.
2. B. C. Grau., C. Halaschek-Wiener., and Y. Kazakov., "History matters: Incremental ontology reasoning using modules", in Proc. ISWC/ASWC, Busan, Korea, pp. 183–196, 2007.
3. H. Paulheim and C. Bizer, "Type inference on noisy RDF data," in *Proc. ISWC*, Sydney, NSW, Australia, pp. 510–525 ,2013.
4. D. Lopez, J., M. Sempere., and P. Garcia., "Inference of reversible tree languages", IEEE Trans. Syst., Man, Cybern. B, Cybern, vol. 34, pp. 1658–1665, Aug. 2004.
5. J. Weaver and J. Handler., "Parallel materialization of the finite RDFS closure for hundreds of millions of triples", in Proc. ISWC, Chantilly, VA, USA, pp. 682–697, 2009.

BIOGRAPHY

Ms. Shruthi Bariki is M-Tech student in Computer Science and Technology, STJIT College of Engineering and Technology of VTU University. She has received Bachelor of Engineering degree in 2012 from GEC Huvina Hadagali, India. Her areas of interests are big data and data mining.