IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

Significant Big Data Interpretation Using SAMR Scheduling Algorithm

M. Suriya[1], S. Jeevitha[2]

PG Scholar, Department of CSE, Kalasalingam Institute of Technology, Krishnan Koil, Virudhunagar, India[1]

Assistant Professor, Department of CSE, Kalasalingam Institute of Technology, Krishnan Koil, Virudhunagar, India[2]

Abstract: Big data is a term for information sets that are so extensive or complex that traditional data processing applications are inadequate. Map Reduce concept is used with an incremental and distributed inference method for large-scale ontologies which realizes high-performance analysis and runtime searching, but it does not find out the slow node at execution time. This paper proposes to improve the efficiency of the map reduce scheduling algorithms by using SAMR scheduling technique which uses the factual information and finds the slow node and launches multi tasks. In addition, the usage timing of each user is calculated. Finally, implement and test the effectiveness of the proposed approach on the Hadoop framework. The purpose of this paper is to speed up the query to the user.

Index terms: Big data, SAMR, Map Reduce, ontology reasoning.

I. INTRODUCTION

With a bigger size of phonological Web [2] information and their fast gain, diverse applications have emerged in a advantage of domains such as healthcare and life sciences, business process management, expert systems, e-marketplace, Web service composition, and cloud system management. Resource description framework (RDF) [8] [11] is a general framework for reporting website metadata, or "data about the data" on the website. It produces interoperability between applications that transfer machine-understandable information on the Web. Traditional centralized logistics methods are not sufficient to process large reasoning. Distributed reasoning methods [10] are thus required to develop the scalability and performance of inference. By encouraging the inclusion of phonological content in web pages, the phonological Web [2] aims at converting the present web, ruled by unstructured and semi-structured script into a "web of data". The main scope of the phonological web [2] is driving the evolution of the current Web by enabling users to find and combine information more easily. Phonological web is based RDF [8] [11], which integrate a variety of applications by using extensible markup language (XML) for pattern and universal resource identifier (URI) naming. Resource description framework (RDF) is a basic representation of ontologies used to express the knowledge on the phonological web. Here, present two concepts, there is Transmission Inference Forest (TIF) and Effective Assertional Triples (EAT).

II. RELATED WORK

A. Antoniou G and Bikakis A [2]
This introduces an implemented defeasible reasoning system (DR-Prolog), which has been tried, assessed and contrasted and existing comparative executions. Through the description of the system, process demonstrated how users can combine the expressive force of a non-monotonic logic (defeasible logic) with the Semantic Web technologies (RDF (S), OWL, Rule-ML) to fabricate applications for the logic and proof layers of the Semantic Web. Entirely describes reasons for conflicts among rules arise naturally on the Semantic Web. To address this issue, we proposed to use defeasible reasoning that is known from the area of knowledge representation, and they had reported on the implementation of a system for defeasible reasoning on the Web. The proposed system is a Prolog-based, bolsters Rule-ML syntax, and can reason with monotonic and non-monotonic rules, RDF facts and RDFS and OWL ontologies.

B. Grau B.C, Halaschek-Wiener C, Kazakov Y [5]
This paper proposed a technique for incremental ontology reasoning that reuses the outcomes got from past calculations. This depends on the thought of a module and can be connected to subjective questions against ontologies communicated in OWL. Here mainly focus on a particular kind of modules that display an arrangement of convincing properties and apply our technique to incremental order of OWL ontologies. It did not depends on a particular reasoned or reasoning method. Here connected to incremental classification of OWL ontologies. For ontology development, it is desirable to re-classify the ontology after a little number of changes. In this situation, our outcomes are exceptionally encouraging. Incremental classification using modules is about ongoing for all ontologies and therefore the reasoned could be working straightforwardly to the client out of sight without backing off the altering of the ontology.

C. Paulheim H and Bizer C [9]
A RDF information base consists of an A-box, i.e., the definition of instances and the relations that hold between

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

them, and a T-box, i.e., a pattern or ontology. The SD-Type approach proposed in this paper exploits links between instances to infer their types using weighted voting. Expecting that specific relations occur only with particular types, they can heuristically accept that an instance should have certain types that it is associated with different instances through specific relations. For every property in a dataset, there is a characteristic distribution of types for both the subject and the object. They had discussed the SD-Type approach for heuristically completing types in expansive, cross-domain databases, based on statistical distributions. Unlike traditional reasoning, this methodology was capable of dealing with noisy data as well as faulty schemas or unforeseen usage of schemas. This procedure can be connected to for all intents and purposes any cross-domain dataset.

D. Lopez D, Sempere J.M, García P [7]
The deduction of tree languages is related to the derivation of context-free string languages using a structural sample, however the improvement of particular tree language learning algorithms should open new possibilities for the characterization of sub-classes of the context-free languages. The two classes of tree languages are characterized, some properties concerning these classes are proven, and they are also studied in relation to other well-known tree language classes. The first class of tree languages is obtained by extension of the notion of reversibility from string languages to tree languages.Here demonstrated that this class contains a few classes of tree dialects and we propose a calculation which takes in the class in polynomial time complexity with respect to the size of the training sample set. The class of reversible tree dialects could be seen as a capacity recognizable dialect, and accordingly it is conceivable to utilize the plan of all.

III. THEORETICAL ANALYSIS

A. Project Scope
The scope of the project to find out the slow node at execution time. A SAMR scheduling algorithm, which improve the performance of resources through dynamic arrangements of resource allocation and reduce the usage of timing. More complicated queries can be decomposed into basic query type and through joining or merge the result the final query result.

B. Problem Statement
In preceding method, the Map Reduce [3] approach is used with an incremental and distributed inference method [1] for large-scale ontologies which realizes high-performance analysis and runtime searching, but it does not find out the slow node at run time. so, it takes more time to execute the process at a delay node occurs.

C. Proposed System
The goal of this project is to find out the delay node and calculate the usage of time. In this paper provide SAMR scheduling algorithm is used for incremental and distributed inference method for large-scale ontologism, which improve the performance of resources through dynamic arrangements of resource allocation.

D. Self Adaptive Map Reduces Algorithm
The SAMR technique uses the actual information that is being stored in every hub and utilizing that data it finds the real slow tasks. Then it maps the slow function and reduces the slow function.

i. Algorithm steps:
- Step 1: input: Key/Value pairs.
- Step 2: output: Statistical results.
- Step 3: read historical information.
- Step 4: tune parameters using proposed k-means clustering.
- Step 5: Find slow tasks.
- Step 6: Find slow task trackers.
- Step 7: Launch back up tasks.
- Step8: Using the results update the historical information.

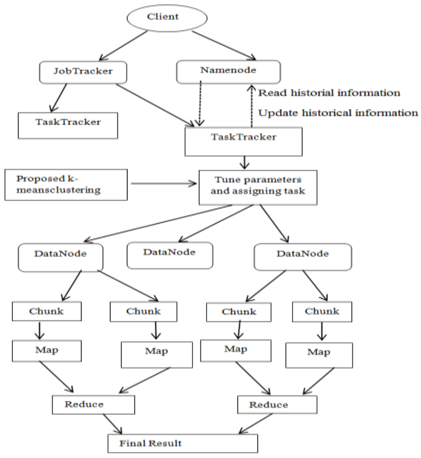

Fig. 1 self adaptive map reduce process

E. K-Means Algorithm
In this paper, use the k-means clustering technique to tune the parameters in the recorded data and finding the moderate errands precisely. The proposed K-means algorithm can solve even the most difficult clustering issues. It requires the quantity of groups that we are going to use in our technique. The algorithm finds k centroids, one for each cluster. Depending on the location of the centroid the result will vary. During the map phase it finds the M1 temporary value and utilizing this worth it finds as a part of the groups which one is closest to the M1 value. Similarly, in the reduction phase it finds the R1 temporary value and using this value it finds in the clusters which one is nearest to the R1 value. Based on the result the centroid area is changed and qualities are recalculated once more.

i. Algorithm steps:
- Step 1: Input: D-set of n data nodes, n-number of data nodes, C-set of k centroids, k-number of clusters.
- Step 2: Output: A-set of k clusters.
- Step 3: Compute distance between each data nodes to all centroids.
- Step 4: For each Di find the closest Ci.
- Step 5: Add Di to A.
- Step 6: Remove Di from D.
- Step 7: Repeat for all Di……Dn and Ci...Ck.

 DOI 10.17148/IJARCCE.2016.55230

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

ii. Algorithm Explanation
function Map is input: integer K1 between 1 and 1100, representing a batch of 1 million social.person records
for eachsocial.person record in the K1 batch do
let Y be the person's age
let N be the number of contacts the person has
k-means clustering aims to partition n observations into k clusters in which each observation belongs to produce one output record (Y,(N,1))
repeat
function cache is searching and avoid the map phase
let K be the intermediate file having the search content
fetch(K)
output
else
map()
function Reduce is
input: age (in years) Y
for each input record (Y,(N,C)) do
Accumulate in S the sum of N*C
Accumulate in C_{new} the sum of C
repeat
let A be S/C_{new}
produce one output record $(Y,(A,C_{new}))$
end function

IV. SYSTEM IMPLEMENTATION

A. Architectural Diagram

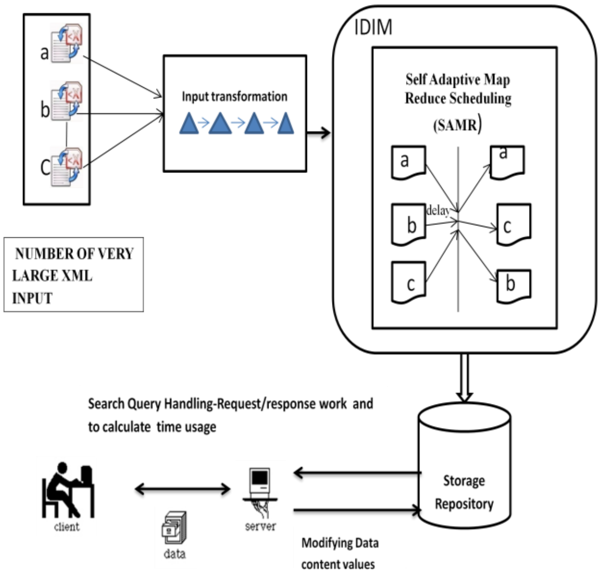

Fig. 1 Architecture Diagram for proposed system

First to relocate the amount of the required files to IDIM. The SAMR map reduce scheduling technique is being developed which uses the factual information and find the slow nodes and launches backup tasks. The historical information is stored in each node in XML format. It adjusts time, weight of each stage of the map and reduces tasks according to the factual information respectively. It decreases the execution time of map reduce job and improve the overall map reduce performance in the heterogeneous environment.

V. RESULT

The proposed k-means clustering algorithm to improve the performance of the Self-adaptive Map Reduce scheduling algorithm. The proposed k-means clustering algorithm to find the nearest distance between each data nodes and each centric. Using this result, it updates the historical information in the name node and find the accurate slow tasks, launch backup tasks and assign tasks to each task trackers. This proposed method takes less amount of computation time.

VI. CONCLUSION

In this paper, proposed a technique to enhance the capacity of the map reduce scheduling algorithms. It works better than an existing map reduces scheduling algorithms by taking less measure of calculation and gives high exactness. Use the proposed k-means clustering algorithm together with the Self-Adaptive Map Reduce (SAMR) algorithm. However, this method works well it can assign only one task to each data node. In the future to improve its capability by assigning more number of undertakings to the data nodes.

REFERENCES

[1] Keman Huang, Jianqiang Li, and MengChu Zhou, Jan- 2015," An Incremental and Distributed Inference Method for Large-Scale Ontologies Based on MapReduce Paradigm".
[2] Antoniou G and Bikakis A, 2007-"DR-Prolog: A system for defeasible reasoning with rules and ontologies on the Semantic Web," IEEE Trans. Knowl. Data Eng., vol. 19, no. 2, pp. 233–245.
[3] Billion Triples Challenge 2012 Dataset [Online]. Available: http://km.aifb.kit.edu/projects/btc-2012/
[4] Dean J and Ghemawat S, 2008-"MapReduce: Simplified data processing on large clusters," Commun. ACM, vol. 51, no. 1, pp. 107–113, 2008.
[5] Grau, B.C, Halaschek-Wiener C, and Kazakov Y, 2007- "History matters: Incremental ontology reasoning using modules," in Proc. ISWC/ASWC, Busan, Korea, pp. 183–196.
[6] Hadoop [Online]. Available: http://hadoop.apache.org/
[7] Lopez D, Sempere J.M, and García P, 2004- "Inference of reversible tree languages," IEE E Trans. Syst., Man, Cybern. B, Cybern.,vol. 34, no. 4, pp. 1658–1665.
[8] Milea V, Frasincar F, and Kaymak U, 2012 "tOWL: A temporal web ontol-ogylanguage," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 42, no. 1, pp. 268–281.
[9] Paulheim H and Bizer C,2013- "Type inference on noisy RDF data," in Proc. ISWC, Sydney, NSW, Australia, pp. 510–525.
[10] Schlicht A and Stuckenschmidt H, 2011-"Map Resolve," in Proc. 5th Int. Conf. RR, Galway, Ireland, pp. 294–299.
[11] Urbani J, Kotoulas S, Oren E, and Harmelen F, 2009- "Scalable distributed reasoning using map reduce," in Proc. 8th Int. Semantic Web Conf., Chantilly, VA, USA, Oct. 2009, pp. 634–649.
[12] Weaver J and Hendler J,2009 "Parallel materialization of the finite RDFS closure for hundreds of millions of triples," in proc.

BIOGRAPHY

M. Suriya received the B.E degree in computer science engineering from Kalasalingam Institute of Technology, Krishnan Koil 2014, M.E in Computer Science Engineering from Kalasalingam Institute of Technology, Krishnan Koil 2016. Her current research in Big Data.