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

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

Load Re-Balancing using Linear Programming for HDFS in Clouds

Pravin S Patil[1], Prof R.R. Badre[2]

P.G. Student, Department of Computer Engineering, MIT Academy of Engineering, Alandi, Maharashtra, India[1]

Associate Professor, Department of Computer Engineering, MIT Academy of Engineering, Alandi, Maharashtra, India[2]

Abstract: A distributed file system often forms a central building block for cloud computing based on Map-Reduce Programming paradigm. The nodes in file system simultaneously serve computers in network and perform storage operations. During processing the file gets partitioned into multiple number of chunks so that Map Reduce task can be performed in parallel over the nodes. Often nodes fail when they upgrade, replace or add. File operation such as create, delete and append results in load imbalance in DFS. Uniform distribution of file chunks gets bit difficult in distributed file system. The centralized load balancing technique includes central load balancer which emerges with bottleneck problem. Proposed system focuses on load rebalancing problem in distributed file system and aims to design a load rebalancing algorithm with linear technique approach to allocate file chunks appropriately to system considering low cost, less execution time. The system will also include Hadoop Distributed File System for maintaining logs and data sets.

Keywords: Cloud Computing, DFS (Distributed file system) Load Balancing, Linear Programming, Hadoop Distributed File System.

I. INTRODUCTION

In cloud computing the computers connect using communication network. The internet is mandatory to perform cloud operations such as delete, create, append and replacement according to cloud notation. The cloud users share information and resources using cloud operations. The cloud is popular for features such as Scalability, On Demand Service, Pricing, Quality of Service etc. For data storage purpose the technologies such as Map-Reduce distributed file systems, virtualization are incorporated in Cloud.

The Distributed file system technology uses technique of forming chunks on cloud computing application based on the Map-Reduce programming paradigm which follows master-slave architecture, i.e Master which act like Name node and Slave which act like Data node. Master/Server takes large problem, divides it into sub problem and assigns it to worker node i.e. to multiple slaves to solve problem individually. Map-Reduce programming technique in Distributed File System divides a large file into number of chunks and allocates each chunk to separate node to perform Map-Reduce function parallel over each node. e.g Consider word count application task which identifies occurrences of each distinct word in large file. The submitted large file is divided in fixed-size chunks or parts and assigned to different cloud storage node. Large file is divided in fixed-size chunks or parts by system and assigned them to different cloud storage nodes. The storage node does its task by calculating occurrences of each distinct word by scanning and parsing its own chunk. Server/Master collects result from each storage node to calculate the final result.

In Distributed file system the load of each storage node is directly proportional to number of file chunks that node consists. In cloud systems increase in storage and network is the main issue for balancing load in large scale distributed systems. There is need to balance the load over multiple node to improve system performance, resource utilization, response time and stability. The storage nodes, in cloud, number of files and assesses to that file increases then the central node (master in Map-Reduce) becomes bottleneck. The load imbalance problem can be solved by designing load balancing algorithm, in which storage nodes are structured over network based on the distributed hash table (DHT); each file chunk having rapid key lookup in DHTs, in that unique identifier is assign to each file chunk.

The proposed system aims to reduce the movement cost which is caused by load imbalance problem due to nodes used to maximize the network bandwidth. In proposed system the Submitter client application will submit the task to server. Post receiving of the task, server will distribute the task to volunteer clients/storage nodes. Algorithm like Linear Programming may be used for load distribution based on current load that will submitted to volunteer clients/storage nodes. The volunteer clients/ storage nodes will complete the process and send the response to the server. Server intern will collect the result and send the reply back to Submitter client application.

II. RELATED WORK

An Load Rebalancing for Distributed File Systems in Clouds [1] paper focuses on fully distributed load

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

rebalancing algorithm to cope with the load imbalance problem. In production system the algorithm is compared against centralized approach and a competing distributed solution presented in the literature. In future proposed work will be implemented in Hadoop distributed file system incorporating in cluster environment.

A New Approach to Improve Load Balancing for Increasing Fault Tolerance and Decreasing Energy Consumption in Cloud Computing[2] The aim of the system is to present a novel strategy to improve load balancing for increasing fault tolerance and reducing energy consumption via ranking the tasks and virtual machines in cloud computing by fuzzy method.

Analytical Literature Survey on Existing Load Balancing Schemes in Cloud Computing [3] this system focuses about balancing the load in cloud environment, need of load balancing, existing literature on load balancing algorithms, and widely used performance metrics for load balancing and detailed analysis on the algorithms taken in the literature.

A secured load balancing architecture for cloud computing based on multiple clusters[4] The paper focuses on how the system address the subject of load balancing in cloud computing present a semi centralized and multi cluster architecture. This proposed approach mainly ensures a better overall performance with efficient load balancing, the continuous availability and a security aspect.

Efficient Utilization of Virtual Machines in Cloud Computing using Synchronized Throttled Load Balancing [5] The system propose an algorithm that focuses on load balancing to reduce the situation of overload or under load on virtual machines that leads to improve the performance of cloud substantially. Comparative analysis is being done with the help of Cloud Analyst tool.

Cluster Based Load Balancing in Cloud Computing [6] The system specifies that Cluster based load balancing works well in heterogeneous nodes environment which considers resource specific demands of the tasks and reduces scanning overhead by dividing the machines into clusters. The algorithm used gives better results in terms of waiting time, execution time, turnaround time and throughput as compared to existing throttled and modified throttled algorithms.

Load Balancing in Cloud Computing Using Dynamic Load Management Algorithm [7] The system presents a dynamic load management algorithm proposed for distribution of the entire incoming request among the virtual machines effectively. The performance is simulated by using Cloud Analyst simulator based on various parameters like data processing time and response time etc. After comparing results with VM Assign algorithm the present algorithm distributes the load uniformly among server through efficient usage of resources uniformly.

Load Balancing in Cloud Computing Environment Based on An Improved Particle Swarm Optimization [8] Proposed system focuses on improved particle algorithm to achieve resource load balancing optimization in the cloud environment. The mechanism collects the characteristics of complex networks into consideration to establish a corresponding resource-task allocation model. The simulated experiments showed that this model can improve the load balancing and resource utilization in the cloud.

Load Balancing Job Assignment for Cluster-Based Cloud computing [9] The proposed work combines the characteristics of cloud with Grid Computing to achieve load-balancing mechanism. The system contributes to adopt queuing model to calculate the round-trip time and nodal processing to evaluate the end-to-end delay.

Cloud Light Weight: a New Solution for Load Balancing in Cloud Computing [10] The system focuses a new load balancing method, as Cloud Light Weight (CLW), which balances the Virtual Machines'(VM) workload in cloud computing DC, but it also assures QoS for users.

III. DRAWBACKS OF EXISTING SYSTEM

In distributed file system, the central node becomes a performance bottleneck because of increase in storage nodes, files and number of accesses to file. The dependence of central node emerge with bottleneck problem results in load imbalance issue which possess heavy load.

IV. PROPOSED SYSTEM

Proposed system focuses on load rebalancing problem in distributed file system and aims to design a load rebalancing algorithm. This will result in allocation of file chunks appropriately to system considering low cost, less execution time, incorporating cloud with linear technique approach to obtain fine optimized results including Hadoop Distributed File System for maintaining logs and data sets. The system will also improve Resource Utilization; perform Optimized Scheduling so that end user can get Cloud Cost Justification.

V. SYSTEM ARCHITECTURE

The system will consist of three different entities like Submitter Client Application (Submits task to server), Cloud Server (For load balancing) and Ad-Hoc Volunteer client (Process the task and submit response to server). Fig 1 shows Architecture of Proposed System.

Client/Submitter Application (Which will Submits task to server)

- The submitter application will submit Image Batch for Processing to server.
- The task will be time consuming which actually takes time to execute.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

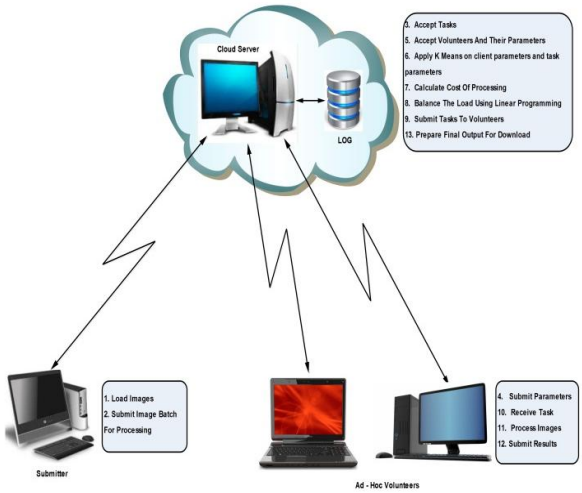

Fig 1: Architecture of Proposed System

Cloud Server (For load balancing)

- Firstly server will accept task
- It will accept volunteers and their parameters.
- It will apply K-means on client parameters and task parameters.
- It will calculate cost of processing.
- The load will be balanced using Linear Programming
- The task will be submitted to the volunteers.
- Lastly server will prepare final output for Download.

Volunteer client (Process the task and submit response to server)

- Ad –Hoc Volunteer will submit parameters to server.
- It will be responsible to receive task.
- The image will be processed.
- The result will be submitted back to the Cloud Server.

The Fig.2 shows detailed Architecture of the system.

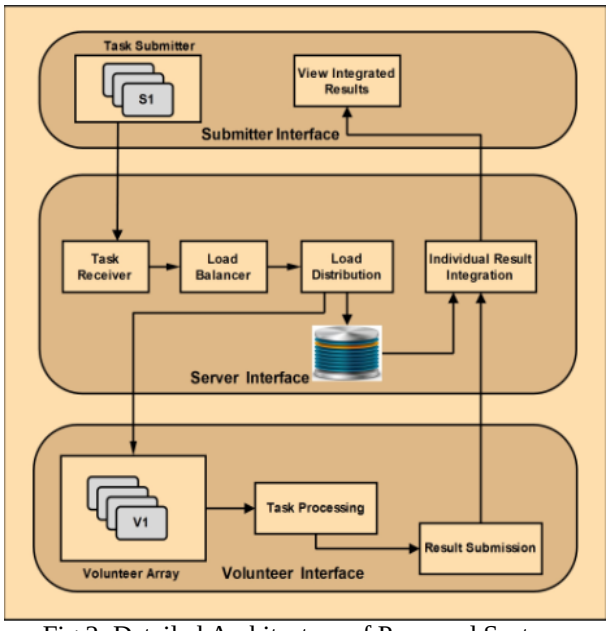

Fig.2: Detailed Architecture of Proposed System

Task Submitter

• The task submitter is responsible to submit the task to server. There can be any 'n' number of submitter. The cloud is capable of executing all of them. Hence there is no limitation on number of submitters.

Task Receiver

• Task receiver is responsible for receiving and forwarding task to load balancer. In proposed system the task shall be an image on which basic reprocessing will be performed. Lets consider I←{I1,I2,I3…In}, Where I is set of image on which pre-processing shall be performed.

Load Balancer

• After receiving task the load balancer divides it in chunks and distribute them among the number of volunteer. If I1 is an input image of size 250 X 250 where we have total number of bytes as 62,500 and we have for example 3 volunteers so image shall be breaked into chunks as I1C1, I1C2, I1C3. After this a load balancing algorithmic equation shall be applied based on which image will be passed to load distributor.

Load Distribution Module

• Load distribution is responsible to forward the image to the volunteer clients based on parameters calculated from load balancer.

Volunteer Array

• There may be any number of 'n' volunteer in the network. The task received from load distribution shall be processed here. For example assume the task in pre-processing of image in Gray Scale. The output image will be send back to server via Result Submission.

Individual Result Integration

• The braked image in chunks is to be integrated together.

VI. SYSTEM FLOW

The proposed system consists of the following modules:

A. System Initialization

B. Client/Submitter Application
1) Login
2) Generate_Code
3) Submit to Server
4) View Result

C. Cloud Server/Cloud Service Provider
1) Connect
2) Load Balance
3) Linear Programming
4) Relay Task
5) Process task
6) Submit Task

D. Volunteer client/Process task
1) Data Storage

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

Data Processing

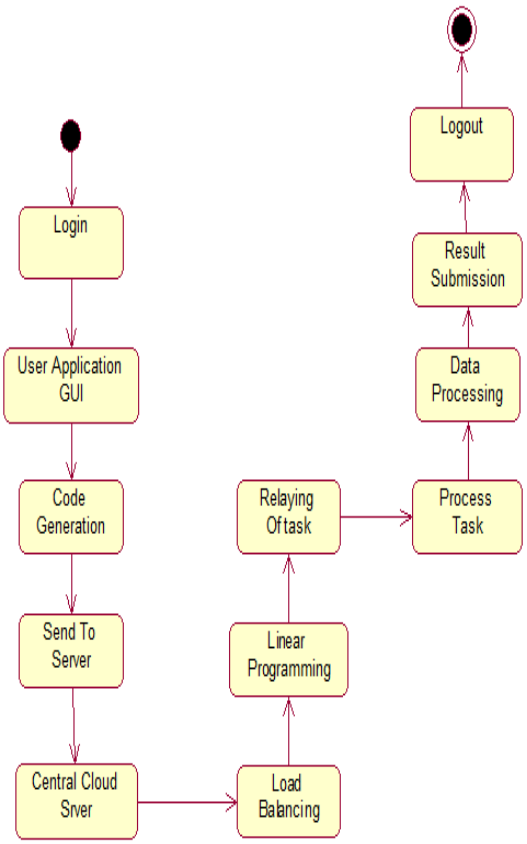

Fig.3: Flow of Proposed System

The system can be implemented by above mentioned modules. Once system is initialized the Client/Submitter Application is responsible for client registration. Once registration process is complete the client can login to system and can submit image task for processing to the server.The Cloud Server/Cloud Service Provider is responsible to establish connection , accept task and to collect volunteer parameters. This can be done by applying K-means on client and task parameters. This will result in calculating cost of processing. The load distribution among Volunteer client can be done on the basis of linear programming. The image task will be submitted to the volunteers for processing. Server will be responsisble to submit task back to client i.e preparing final output for download from client side. The volunteer client will process task and maintain data storage.

VII. CONCLUSION

The proposed system will resolve load re-balancing problem in distributed file systems specialized for large-scale, dynamic and data-intensive clouds by allocating the chunks of files as uniformly as possible among the nodes such that no node manages an excessive number of chunks. Particularly, system will have linear algorithm/technique which will exhibit a fast convergence rate along with fine optimized results and system accuracy. The system will be in-corporating Hadoop Distributed File System for maintaining logs and data sets.

REFERENCES

[1] Hung-Chang Hsiao, Haiying Shen,"Load Rebalancing for Distributed File Systems in Clouds" IEEE Transactions On Parallel And Distributed Systems, VOL. 24, NO. 5, YEAR 2013.
[2] Ali Moghtadaeipour, Reza Tavoli,"A New Approach to Improve Load Balancing for Increasing Fault Tolerance and Decreasing Energy Consumption in Cloud Computing" , IEEE,2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI)
[3] Garima Rastogi, Dr Rama Sushil" Analytical Literature Survey on Existing Load Balancing Schemes in Cloud Computing" ,IEEE, 2015 International Conference on Green Computing and Internet of Things (ICGCIoT)
[4] Mohamed Belkhouraf, Ali Kartit and Hassan Ouahmane, Mohamed Belkhouraf, Ali Kartit and Hassan Ouahmane" A secured load balancing architecture for cloud computing based on multiple clusters" IEEE, 2015 International Conference on Cloud Technologies and Applications (CloudTech)
[5] Shikha Garg_, Rakesh Kumar Dwivediy and Himanshu Chauhanz" Efficient Utilization of Virtual Machines in Cloud Computing using Synchronized Throttled Load Balancing" IEEE, 2015 1st International Conference on Next Generation Computing Technologies (NGCT)
[6] Surbhi Kapoor, Dr. Chetna Dabas" Cluster Based Load Balancing in Cloud Computing" IEEE, Eighth International Conference on Contemporary Computing (IC3 2015)
[7] Reena Pan war, Prof. Dr. Bhawna Mallick" Load Balancing in Cloud Computing Using Dynamic Load Management Algorithm" IEEE, 2015 International Conference on Green Computing and Internet of Things (ICGCIoT)
[8] Kai Pan, Jiaqi Chen" Load Balancing in Cloud Computing Environment Based on An Improved Particle Swarm Optimization",IEEE, 2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS)
[9] Yean-Fu Wen and Chih-Lung Chang" Load Balancing Job Assignment for Cluster-Based Cloud Computing" ,IEEE, 2014 Sixth International Conference on Ubiquitous and Future Networks (ICUFN)
[10] Mohammadreza Mesbahi, Amir Masoud Rahmani, Anthony Theodore Chronopoulos" Cloud Light Weight: a New Solution for Load Balancing in Cloud Computing" IEEE, 2014 International Conference on Data Science & Engineering (ICDSE)

BIOGRAPHIES

Mr. Pravin S Patil, M.E Student, Dept of Comp Engg, MIT, AOE, Alandi (D), Pune. Received B.E. degree in Computer Engineering from North Maharashtra University, Jalgaon, India, in 2007. Interesting Area is Cloud Computing.

Prof. R.R. Badre received her M.E. degree in Computer Science and Engineering from Shivaji University, Kolhapur. She is currently working as an Associate Professor in the Department of Computer Engineering in MIT Academy of Engineering, Pune. She has published more than 8 papers in both International journals and conferences. She is also a member in Indian Society of Technical Education.