ENERGY-EFFICIENT AND COST-EFFECTIVE RESOURCE PROVISIONING FRAMEWORK FOR MAP REduce WORKLADOFS USING DCC ALGORITHM

G.Anuprabavathi¹, R.Rajmohan², J.Nulyn Punitha³, D.Dinagaran⁴, S.G.Sandhya⁵
Research Scholar, Assistant Professor¹,²,³,⁴,⁵, Department of Computer Science & Engineering, IFET College of Engineering, Villupuram, India
anuprabavathig@gmail.com¹, rimohan89@gmail.com², mailnulyn@gmail.com³, ddinagaran@gmail.com⁴, sgsandhyadhas@gmail⁵

Abstract: Resource Provisioning is an important and challenging problem in the large-scale distributed systems such as Cloud computing environments. The growing trend in cloud computing is the combination of Big Data and Big Data analytics that has been driven with rapid evolution of data centre technologies towards more cost-effective solutions. Map Reduce profiling is used to automatically create the best cluster configuration for the jobs scheduling. The existing models provide a globally efficient resource allocation scheme that significantly reduces the resource usage cost in the cloud resource optimization for the jobs. But the energy efficiency is at critical concern due to evolve of Map reduce workloads. The proposed Energy Efficient and Cost effective (EECE) resource management framework aim to minimize the infrastructure cost in the data centre and energy conservation for cloud clusters. The proposed system is to reduce the cost for allocating the resources using the virtual clusters globally. The proposed system reduces the cost by eliminating the reconfiguration based approach. The proposed approach is based on integer partitioning approach which selects optimal nodes in a dynamic cloud environment to configure a cluster for running Map Reduce jobs.

The proposed approach is cost optimized, adheres to global resource utilization and provides high performance to the clients. Experiment results shows that our algorithm performs better than the existing scheduling algorithms in terms of energy utilization and minimal cost.

Keywords: Map reduce, Energy efficient, Cost effective, DCC algorithm

I. INTRODUCTION

The most popular approach is using Map Reduce for big data analytics. Existing per-job services that require VMs to be created a fresh for each submitted job [6], EECE deals with such interactive workloads using a secure instant VM allocation scheme that minimizes the job latency. Existing cloud solutions are largely optimized based on per-job and per- customer optimization which leads to poor resource utilization and higher cost [5]. The first operational model (immediate execution) is a completely customer managed model where each job and its resources are specified by the customer on a per-job basis and the cloud provider [9] only ensures that the requested resources are provisioned upon job arrival.

This model has the lowest rewards since there is lack of global optimization across jobs as well as other drawbacks discussed earlier. It is difficult to apply access control enforcement while the workflow is being executed, if the access rights of jobs change dynamically [7]. The leased clusters are under-utilized for a large fraction of the time leading to higher costs. The techniques result in poor utilization lead to higher cost [8].

The proposed Energy Efficient and Cost effective (EECE) resource management framework aim to minimize the infrastructure cost in the data centre and energy conservation for cloud clusters. The proposed system is to reduce the cost for allocating the resources using the virtual clusters globally. The proposed system reduces the cost by eliminating the reconfiguration based approach. The proposed cluster configuration is based on integer partitioning approach which selects optimal nodes in a dynamic cloud environment to configure a cluster for running Map Reduce jobs.

The proposed approach is cost optimized, adheres to global resource utilization and provides high performance to the clients. Experiment results shows that our algorithm performs better than the existing scheduling algorithms in terms of energy utilization and minimal cost.
II. RELATED WORK

Cost effective in map reduce [1] describes the new cloud managed map reduce model that is designed to provide cost effective solution to effectively handle map reduce workloads. Our EECE system leverages the map reduces profiling that automatically creates the best cluster configuration for the jobs. EECE resource management techniques include cost-aware resource provisioning, vm aware scheduling and online virtual machine reconfiguration. Advantages of this cloud managed model are lower resource usage cost and automatically creates best cluster configuration for the interactive workloads. Cost for reconfiguration of clusters is the disadvantage of this map reduce model.

IEDA process [3] describes an intelligent economic approach for dynamic resource allocation (IEDA) with improved double auction protocol. It is used to enable various kinds of resources traded among multiple consumers and providers at the same time enable task partitioning among multiple providers. A price matching formation is proposed and it consists of a Back Propagation Neural Network based price prediction algorithm and price matching algorithm. In each round of the auction, Consumers and providers submit their bidding and asking prices. Both price related factors and non price related factors can significantly influence the offers. Not only the instant market status but also historical market experience also affects the pricing decisions. At the end of the auction, which provider offers the demanded service to which customer based on the eligible transaction relationship at the same time Winner determination algorithm is needed to those participants who can not only bring economic efficiency but also good reputation are chosen as winners. The advantages with this scheme are human intervention is not needed, improved quality of service. It is a trustful auction.

Dynamic MR [4] describes slot allocation system called DynamicMR that can improve the performance of Map Reduce workloads. It is a technique called Dynamic Hadoop Slot Allocation by keeping the slot-based model. It relaxes the slot allocation constraint to allow slots to be reallocated to either map or reduce tasks depending on their needs.

Dynamic MR System has the following optimization techniques.1) Dynamic Hadoop Slot Allocation (DHSA) 2) Speculative Execution Performance Balancing (SEPB) 3) Slot Pre Scheduling technique.

III. PROPOSED SYSTEM

The proposed system is able to do the cluster configuration without the reconfiguration component. Reconfiguration stage is eliminated using dynamic cluster configuration algorithm.

Cluster Configuration algorithm that selects optimal nodes in a dynamic cloud environment to configure a cluster for running Map Reduce jobs. The algorithm is cost optimized, adheres to global resource utilization and provides high performance to the clients. Use of an integer partitioning based dynamic clustering algorithm eliminates the concept of VM pools and it is able to select the right set of nodes for cluster configuration. When selecting the clusters for performing the Map reduce jobs, the node which is having the lowest CPU utilization will get more priority than other. Hence the CPU utilization is also globally improved.

I. JOB PROFILING AND ANALYSIS

Users simply submit their jobs (or composite job workflows) and specify the required service quality with the respective deadline. Profile and analysis phase develops a performance model for the job in order to be able to generate predictions for its performance later on. When making scheduling decisions, performance predictions in terms of execution time are made based on the input dataset size, VM types, cluster sizes and job parameters. The fig.1 describes the architecture of the proposed system.
The output of the profiling phase is the optimal number of nodes that can perform the job. This optimal number of nodes is the basis for the cluster configuration for the Map Reduce job. Profile and analyze service is shown in figure 2.

2. INTEGER PARTITIONING APPROACH

The number of nodes is processed by using the integer partitioning algorithm. An integer partitioning algorithm that generates all possible combination of numbers which is result to the output of the profiling phase. The nodes are hence allocated to configure a cluster. In this approach, initially the switch has the number of active nodes available in the network. Within these available nodes only cluster configuration is implemented.

3. VM ALLOCATION AND SCHEDULING

Once the optimal numbers of nodes are chosen, the node which is active inside the switch is chosen for cluster configuration. There will be so many nodes to form a cluster. In that nodes which have the lowest CPU utilization will get more priority than other nodes. So that VM is allocated for the Map Reduce jobs. The Job scheduler at the cloud provider is an integral component of the proposed system. Where existing Map Reduce services simply provision customer-specified VMs to execute the job, VM-aware scheduler (Section 3.1) is faced with the challenge of scheduling jobs among available VM pools while minimizing global cloud resource usage. Therefore, carefully executing jobs in the best VM type and cluster size among the available VM pools becomes a crucial factor for performance. The scheduler has knowledge of the relative performance of the jobs across different cluster configurations from the predictions obtained from the profile and analyzes service and uses it to obtain global resource optimization. The goal of VM-aware scheduling is to schedule all jobs within available VM pools to meet their deadlines while minimizing the overall resource usage in the data center reducing this total infrastructure cost. As jobs are incrementally submitted to the cloud, scheduling requires an online algorithm that can place one job at a time on an infrastructure already executing some jobs. To better understand the complexity of the problem, we first analyze an offline version which leads us to the design of an online scheduler as shown in Fig.3.
4. CLUSTER CONFIGURATION FOR MAP REDUCE OPERATION

Once VM allocated, cluster configuration has made to do the Map Reduce job. Map reduce job is performed using Hadoop Map Reduce framework.

**Dynamic Cluster Configuration algorithm**

**Input:** Nodes count  
**Output:** Cluster List  
**Require:** Resource Status [1..n] = 3,  
Node List [1..n* m] = 3, AvailableNodes = n * m,  
allSelectedResources [1..n] = 0  

if number of Nodes > Available Nodes then  
Print “Cluster allocation is impossible due to insufficient Resource”  
Else  
Get the permutation free in the mapping list while mapping List != NULL do  
break;  
Identify the nodes.  
Resource, change the colour accordingly  
For allSelectedResource do  
for i = 1 to 4 do  
if NodeList[Resource Number*4 + i] = 1 then  
NodeList[ResourceNumber*4 + i] = 1  
end if  
end for  
allSelectedResources [next]  
end for  
Call this function for all request.  
end while  
end if

IV. EXPERIMENTAL DESIGN

We first present the experimental evaluation of our system by comparing with the existing techniques for various experimental conditions determined by distribution of the job deadlines, size of the Map Reduce jobs, and number of servers in the system and the amount of prediction error in the profile and analyze process. By default, we use a composite workload consisting of equal proportion of jobs of three different categories: small jobs, medium jobs and large jobs. Small jobs read 100 MB of data, whereas medium jobs and large jobs read 500 MB and 1 GB of input data respectively.

The evaluation uses 10,500 jobs arriving within a period of 100 minutes. Each of the arrived job represents one of the 50 profiled jobs with input data size ranging from 100 MB to 1 GB based on the job size category. By default, we assume that jobs run for the same amount of time predicted in the profile and analyze process; however, we dedicate a separate set of experiments to study the performance of the techniques when such predictions are erroneous. The data loading time is computed by assuming a network throughput of 25 MBps per VM from the storage server and the Hadoop startup time is taken as 10 sec.

V. RESULTS AND DISCUSSION

We find that provisioning dedicated clusters for each customer results in a lot of resources as dedicated clusters are based on the peak requirements of each customer and therefore the resources are under-utilized. On the other hand, per-job cluster services require lower number of servers (fig.4) as these resources are shared among the

**Fig.3 Pool of Instances**

Pools of small instance  
Pools of extra large instance  
Pools of large instance
customers. However, the EECE approach in Fig.4 has a much lower resource.

The cost trend shown in Fig.7 also shows that the techniques that require fewer servers result in lower per-job cost.

In whole Our EECE system outperforms the Cura and Starfish framework in terms of Job allocation cost, Job utilization and response time.

VI. CONCLUSION

This paper presents a new MapReduce cloud service model, EECE, for data analytics in the cloud. The existing cloud services for MapReduce are inadequate and inefficient for production workloads. The EECE automatically creates the best cluster configuration for the jobs using MapReduce profiling and leverages deadline-awareness which, by global resource allocation efficiently and reduce its costs. EECE also uses a unique secure instant VM allocation technique that ensures fast response time guarantees for short interactive jobs, a significant proportion of modern MapReduce workloads. EECE resource management techniques include cost-aware resource provisioning, VM-aware scheduling and online virtual machine reconfiguration. Our experimental results using jobs profiled from realistic Twitter-like production workload traces show that EECE achieves more than 70 percent reduction in infrastructure cost with 55 percent lower job response times.

REFERENCE:


