GREEN REVOLUTION APPROACH TOWARDS MANAGEMENT OF RESOURCES IN CLOUD BASED ENVIRONMENT

P. Raghuvara pandian[1], M. Lavanya[2]
[1,2] Department of Computer Science and Engineering, Dr.APJ Abdul Kalam Centre for Research, Adhi College of Engineering and Technology, Sankarapuram - 631605, Tamilnadu, India.
raghuvarapandiyan@gmail.co

Abstract

With the emergence of many new data centers around the globe, energy consumption by those data centers has been tremendously increased. Dynamic capacity provisioning is a promising approach for reducing energy consumption by dynamically adjusting the number of active machines to match resource demands. The main problem in dynamic capacity provisioning is considering the heterogeneity of both workload and machine hardware found in the production environment. To address this limitation, workload has been divided into distinct classes with similar characteristics in terms of resource requirements using K-means clustering algorithm. Moreover based on the resource requirement of each task class, numbers of machines in the data centers are dynamically adjusted in order to minimize the total energy consumption.

Keywords: **data centers, Dynamic capacity provisioning, heterogeneity, K-means clustering algorithm.**

I. INTRODUCTION

Cloud computing has risen as a new computing paradigm to bring unparalleled flexibility and access to shared and scalable computing resources [16]. Cloud services are usually implemented in one or more data centers where a large number of servers, storage units, and a telecommunication infrastructure are provisioned [9].

As these data centers raise from hundreds to hundreds of thousands of servers to meet the increasing demand, the energy cost of these large- scale data centers contributes to a major portion of the operating costs.[10]

Cloud computing is computing in which huge groups of remote servers are networked to allow the centralized data storage, and online access to computer services or resources. Clouds can be classified as public, private or hybrid. A data center is a centralized repository for the storage, management, and dissemination of data and information. Typically, a data center is a facility used to house computer systems and related components, such as telecommunication and storage system.

Often times, there are redundant and backup power supplies, redundant data communication connections, environmental controls, and security devices. Reducing energy consumption has become a primary concern for today's data center operators [19]. One promising technique that has received significant attention is Dynamic Capacity Provisioning (DCP). The goal of this technique is to dynamically adjust the number of active machines in a data center in order to decrease energy consumption while meeting the resource requirements of the workloads.

II. LITERATURE SURVEY

With reduction in energy consumption of IDC's. It mainly addresses task heterogeneity of IDC's. It presents a HARMONY(Heterogeneity Aware Resource MONitoring and management sYstem) to minimize energy consumption[12].Datacenters have recently gained important popularity as a cost-effective platform for hosting large-scale service applications. While huge data centers enjoy economies of scale by amortizing long-term capital investments over a huge number of machines, they also incur remarkable energy costs in terms of power distribution and cooling. For instance, it has been reported that energy-related costs account for approximately 12 percent [12] of overall data center expenditures.

For big companies like Google, a 3 percent reduction in energy expenditure can translate to over a million dollars in cost savings. At the same time, governmental agencies persist to implement and regulations to promote energy-proficient computing. As a result, reducing energy utilization has become a primary concern for today's data center operators. In recent years, there has been wide research on improving data center energy efficiency. One hopeful technique that has received important attention is Dynamic Capacity Provisioning (DCP). The objective of this technique is to dynamically adjust the number of active machines in a data center in order to decrease energy utilization while meeting the Service Level Objectives (SLOs) of workloads. In the context of workload scheduling in data centers, a metric of particular significance is scheduling delay, which is the time a request waits in the scheduling queue before it is listed on a machine. In IaaS (Infrastructure as a Service) also heterogeneity is a challenging issue. [10].