"Resource Scheduling algorithm in Cloud Computing" - A Survey

1st Disha Patel,2nd Jasmine Jha
1st PG Student,2nd Asst.Prof.
1st Information Technology
1st L.J.Institute of Engineering & Technology,Sarkhej, Ahmedabad
[1]dish191191@gmail.com, [2]jhajasmine@gmail.com

Abstract - Resource scheduling is one of the key problems of cloud computing, no wonder, the scheduling policy and algorithm affect the performance of the cloud system directly.resource allocation is performed with the objective of minimizing the costs associated with it. The other challenges of resource allocation are meeting customer demands and application requirements In this paper, various resource allocation strategies and their challenges are discussed in detail. It is believed that this paper would benefit both cloud users and researchers in overcoming the challenges faced.

Index Terms–Cloud Computing, Resource Scheduling.

Introduction

Cloud Computing

Cloud computing means the practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer.In cloud computing, the word cloud is used as a metaphor for "the internet" so the phrase *cloud computing* means "a type of Internet-based computing," where different services — such as servers, storage and applications — are delivered to an organization's computers and devices through the Internet.

Three type of cloud

1. Private cloud: Private cloud services are delivered from a business' data center to internal users.

2. Public cloud: A third-party provider delivers the cloud service over the Internet. Public cloud services are sold on-demand, typically by the minute or the hour.Ex: Amazon Web Services (AWS), Microsoft Azure, Google Compute Engine.

3. Hybrid cloud: Hybrid cloud is a combination of public cloud services and on-premises private cloud – with orchestration and automation between the two.

Cloud divided into three broad service categories

Infrastructure as a Service: IaaS provides virtual machines, virtual storage, virtual infrastructure ,and other hardware assets as resources that clients can provision.The IaaS service provider manages all the infrastructure, while the client is responsible for all other aspects of the deployment. This can include the operating system, applications,and user interactions with the system.

Examples of IaaS service providers include:

- Amazon Elastic Compute Cloud (EC2)
- Eucalyptus
- Go Grid

Platform as a Service: PaaS provides virtual machines, operating systems, applications,Services , development frameworks, transactions, and control structures.The client can deploy its applications on the cloud infrastructure or use applications thatwere programmed using languages and tools that are supported by the PaaS service provider.The service provider manages the cloud infrastructure, the operating systems, andthe enabling software. The client is responsible for installing and managing the application that it is deploying.

Examples of PaaS services are:

- Force.com
- Go Grid Cloud Center
- Google App Engine
- Windows Azure Platform

Software as a Service: SaaS is a complete operating environment with applications, management, and the user interface.In the SaaS model, the application is provided to the client through a thin client interface(a browser, usually), and the customer's responsibility

begins and ends with entering andmanaging its data and user interaction. Everything from the application down to the infrastructure is the vendor's responsibility.

Examples of SaaS cloud service providers are:

- GoogleApps
- Oracle On Demand
- SalesForce.com
- SQL Azure

The three different service models taken together have come to be known as the SPI model ofcloud computing. Many other service models have been mentioned: StaaS, Storage as a Service; IdaaS, Identity as a Service CmaaS, Compliance as a Service and so forth. However, the SPI services encompass all the other possibilities.(a browser, usually), and the customer's responsibility begins and ends with entering and managing its data and user interaction. Everything from the application down to the infrastructure is the vendor's responsibility.

Related Work

Resource Scheduling

In cloud computing, Resource Allocation (RA) is the process of assigning available resources to the needed cloud applications over the internet. Resource allocation starves services if the allocation is not managed precisely. Resource provisioning solves that problem by allowing the service providers to manage the resources for each individual module.

The order and time of allocation of resources are also an input for an optimal RAS. RAS should avoid the following criteria as follows:

a) **Resource contention** situation arises when two applications try to access the same resource at the same time.
b) **Scarcity of resources** arises when there are limited resources.
c) **Resource fragmentation** situation arises when the resources are isolated. [There will be enough resources but not able to allocate to the needed application.]
d) **Over-provisioning** of resources arises when the application gets surplus resources than the demanded one.

Resource users (cloud users) estimates of resource demands to complete a job before the estimated time may lead to an over-provisioning of resources. Resource providers' allocation of resources may lead to an under-provisioning of resources.

TABLE I. INPUT PARAMETERS

Parameter	Provider	Customer
Provider Offerings	√	-
Resource Status	√	-
Available Resources	√	-
Application Requirements	-	√
Agreed Contract Between Customer and provider	√	√

The input parameters to RAS and the way of resource allocation vary based on the services, infrastructure and the nature of applications which demand resources.

Different type of algorithm is used for allocate the resources to available task or job. This type of algorithm help us to find the fit resources for available task.

Types of algorithm for Resource scheduling

- Genetic algorithm
- Bee algorithm
- Ant Colony algorithm
- Workflow algorithm
- Load balancing algorithm

Genetic Algorithm:

In this process is repeated until termination condition is reached. Common termination find on following criteria

1) Allocate budget reached.
2) Fitness function is reached

Bee Algorithm

It is a nature inspired algo which tries to track the activities of bee to get their food Following step are perform:

- no of site selected from visited site
- No of best site out of whole set.
- No of bees recruited for the best site.
- No of recruited for other site

Ant colony algorithm
Ant colony algorithm is useful for combinatorial optimization problem. Besed on the behavior of ants seeking a path between their colony and source food.First 1 ant find the food resource and back to their nest.And after Its take the shortest path from their colony to resource.

- First pheromone initialization
- Location of the ant is an entry state
- Then next state will be selected
- Check next state is final state then go state 5, otherwise repeat step 3.
- Pheromone is updating step.
- If stopping criteria is satisfied then stop otherwise go step 2.

Workflow algorithm
Workflow is process that consist of series of step which simplifies the complexity of execution of application.

- A single workflow consist of set of task and each task communicate with each other.
- Workflow scheduling discover resources and allocate suitable task on resources.
- For proper scheduling various algorithm are used.

Load balancing
Optimal use of resources, maximize throughput, minimize response time and avoid overload load balancing is require.

Literature Review

A Resource allocation strategy Based on particle swarm algorithm in cloud computing[1].
A strategy of resource allocation and price adjustment based on particle swarm algorithm is proposed in this paper. According to workload characteristic utility function is designed is designed for evaluate the QoS. According to resource demand from all workload, the resource price are dynamically adjust by the corresponding resource agent in order to obtain maximum profit from each workload.

A Cloud Computing Resource Scheduling Scheme Based on Estimation of Distribution Algorithm.
Resource scheduling is one of the key problems of cloud computing, no wonder, the scheduling policy and algorithm affect the performance of the cloud system directly. In order to improve the utilization of cloud computing resources and keep load balancing, a cloud computing resource scheduling algorithm based on estimation of distribution algorithm is proposed. In this algorithm, the idea of population based incremental learning (PBIL) algorithm is fully used.

In this paper, Author Firstly establish cloud computing resource scheduling algorithm and then objective solution is made by using the PBIL algorithm. And finally show the simulation results show that the PBIL algorithm can take shorter time to complete task and achieve resource load balancing, especially, for the resource scheduling with large-scale task, the advantages are more apparent.

Dynamic Resource Allocation using Virtual Machines for Cloud Computing Environment.
Cloud computing allows business customers to scaleup and down their resource usage based on needs. Many of the touted gains in the cloud model come from resource multiplexingthrough virtualization technology.

In this paper, we present a system that uses virtualization technology to allocate data center resources dynamically based on application demands and support green computing by optimizing the number of servers in use. We introduce the concept of " skewness " to measure the unevenness in the multi-dimensional resource utilization of a server.By minimizing skewness, we can combine different types of workloads nicely and improve the overall utilization of server resources.

Deadline Constrained Cloud Computing Resources Scheduling for Cost Optimization Based on Dynamic Objective Genetic Algorithm.
Cloud computing resources scheduling is significant for executing the workflows in cloud platform because it relates to both the execution time and execution cost.

In this paper, we propose a genetic algorithm (GA) approach to solve this model. In order to tackle with the tight deadline condition, a dynamic objective strategy is further proposed to let GA focus on optimize the execution time objective to meet the deadline constraint when the feasible solution hasn't been obtained. After obtaining a feasible solution, the GA focuses on optimizing the execution cost within the deadline constraint. Therefore, the proposed dynamic objective GA (DOGA) has adaptive ability to the search environment to different objectives.And results show that DOGA can find better solution with smaller cost than PSO does on different scheduling scales and different deadline conditions. DOGA approach is more applicable to be used in commercial activities.

Deadline - Guaranteed Scheduling Algorithm with Improved Resource Utilization for Cloud Computing.
Propose a scheduling algorithm to enhance both deadline guarantee and resource utilization. We modified the conservative backfilling algorithm by utilizing the earliest deadline first(EDF) algorithm and the largest weight first (LWF) algorithm. The proposed algorithm first score all the jobs arrived at the data center(DC) and sort the jobs in ascending order to servehigh priority job first. The proposed algorithm then select the largest possible backfill job as guaranteeing deadline.Simulation results show that the proposed algorithm significantly improve the performance in terms of resource utilization and deadline guarantee.

Conclusion

Cloud computing technology is increasingly being used in enterprises and business markets. And resource scheduling is the key problem in cloud computing. We show the different algorithm for resource scheduling An effective resource allocation strategy is required for achieving user satisfaction and maximizing the profit for cloud service providers. This paper describe the basic scheduling algorithm and their step. Hence this survey paper will hopefully motivate future researchers to come up with smarter and secured optimal resource allocation algorithms and framework to strengthen the cloud computing paradigm.

Reference

1. Niansheng Chen, Xiaoping Fang, Xin Wang" A Cloud Computing Resource Scheduling Scheme Based on Estimation of Distribution Algorithm" IEEE-2014
2. Zong-Gan Chen, Ke-ling Du, Zhi-Hui Zhan (Corresponding Author)lun Zhang "Deadline Constrained Cloud Computing Resources Scheduling for Cost Optimization Based on Dynamic Objective Genetic Algorithm "IEEE-2015
3. SaeMi Shin, Yena Kim and SuKyoung Lee "DeadlineGuaranteed Scheduling Algorithm with Improved Resource Utilization for Clou Computing" IEEE-2015
4. Xie, Yunyun Du, Hongwei Tian"Resource allocation strategy base on Particle swarm algorithm in cloud environment" IEEE-2013
5. Seematai S. Patil, Koganti Bhavani "Dynamic Resource Allocation using Virtual Machines for Cloud Computing EnvironmentI" IJEAT-2014
6. V. Vinothina, Dr.R.Sridaran, Dr.Padmavath Ganapathi"Survey on Resource Allocation Strategies in CloudComputing" *IJACSA-2012.*