International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net

Vol-11 Issue-01 Sept 2016

THE OVERALL MONITORING AND APPLICATION OF A Cloud-Based Data Efficient Management System

Abhishek Kumar

ABSTRACT

When it comes to cloud computing, there aren't many effective algorithms for scientific workflow tasks allocation and scheduling. Most of the ones that exist are more suited to grid computing, where there is a limited amount of computer resources available, but they won't work in the cloud. Users are able to assign and release computed resources on demand and only pay for the resources they really use using elastic computing, as shown in Amazon's Elastic Compute Cloud (EC2), in contrast to grid computing. That being the case, it's safe to presume that resources are endless. The "illusion of infinite resources" describes this aspect of clouds. Despite the obvious advantages of running scientific processes on the Cloud, consumers still don't have enough information to make an informed decision among competing services, especially when such services aim to accomplish competing or even contradictory goals.

KEYWORDS: Workflow application that is data demanding.

1. INTRODUCTION

The term "cloud computing" refers to a method of delivering online services via a network of remote servers hosted by an organization. The services are rented out to customers according to a pay-as-you-go approach, where the user is charged based on how much they use the services [1]. There are three main categories of services that are offered: Cloud computing, software-defined networking,

As seen in figure 1, there is Platform as a service (PaaS) and Software as a service (SaaS).

Figure 1: Cloud Services [2]

Memory, processors, and other physical resources are provided via IaaS. With PaaS, users may build their own apps on the cloud without installing any software on their own computers. Platform as a service (PaaS) offerings:.Net, etc. Current apps, like Facebook, rely on SaaS to function. Installing software on the user's actual system is not something they are concerned about. The software needed to execute these kinds of apps is available on the cloud.

2. DATA INTENSIVE APPLICATION

Data-intensive computing is a class of parallel computing applications which use a data parallel approach to processing large volumes of data typically terabytes or petabytes in size and typically referred to as big data. Computing applications which dedicate most of their execution time to computational requirements are deemed compute-intensive, whereas computing applications which necessitate large volumes of data and devote most of their processing time to I/O and manipulation of data are deemed data-intensive

A data-intensive application workflow deals with the high workloads of data tocontrol

ijise.net

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net

Vol-11 Issue-01 Sept 2016

than its computations of the data. The another meaning of data intensive deals with the transferring the huge amount of data but computational part deals with the processing of the tasks. The transfer of data consumes more time, and also store that data, than the processing of the data. For characterizing the difference between the data-intensive and computer- intensive one aspect is used which is the CCR (Computation to Communication Ratio). The applications with the lower values of this ratio are data intensive applications in nature

RELATED WORK

Allocation and preparing workflow tasks in Cloud has gained popularity in recent times and few algorithms are proposed to deal with this problem

In the authors developed a model that uses particle swarm optimization (PSO) for task-resource mapping to minimize the overall cost of execution such that it completes within deadline that user specifies. To tackle the problem of choosing resource among different cloud providers, a binary integer program is projected in [5], where the objective is to decrease the total infrastructure capacity under budget and load balancing constraints.

In, the authors offered a model for formulation of the generalized federated placement problem and application of this problem to load balancing and consolidation inside a cloud, wherever one cloud can subcontract workloads to partnering clouds to meet peaks in stipulate. They used an Integer Program formulation of the placement program and provide a 2-approximation algorithm. In [7], the authors proposed a binary integer program formulation for cost-optimal scheduling in hybrid IaaS clouds for deadline constrained workloads. In [8], the authors proposed a set of heuristics to cost-efficiently schedule deadline-constrained computational applications on both public cloud providers and private infrastructure. Although, a large amount of these studies consider unbounded number of resources, however, they convert the initial problem (bi-criteria) to constraints problem.

3. SCHEDULING ALGORITHMS

3.1 COST BASED ALGORITHM

In the cost-based approach, we focus only on minimizing the execution and communication costs of using a set of virtual machines incurred by the execution of a given workflow. However, for each obtained feasible solution by using an allocation strategy

ijise.net

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-11 Issue-01 Sept 2016

the overall completion time corresponding is computed. Recall that the intention is to assign tasks to virtual machines respecting the precedence constraints. The cost-based approach is an application allocation and scheduling algorithm for an unbounded number of virtual machines. As mentioned earlier, users can request and obtain sufficient resources at any time. The approach proposed has three most important phases, namely:

- **Tasks sorting phase**

Tasks sorting phase In order to group the tasks that are independent of each other, the given workflow (DAG) is traversed in a top-down fashion to sort tasks at each level. As a result, tasks belonging to the same level do not exchange data and can be executed in parallel (because they are not related by precedence constraints).

- **Resource allocation phase**

In this phase the selection of an ”optimal” virtual machine for each task is decided. In other words, the virtual machine which gives minimum execution and communication costs for a task is selected and the task is assigned to that virtual machine. Following strategies: top- down, bottom-up and mixed exploration and portion strategy.

The top-down strategy consists of starting by the allocation of the initial task (level 1) to the virtual machine which gives minimum execution cost. After this assignment, the graph is traversed in a top-down fashion from level 2 to level L.

The bottom-up strategy consists on starting by the allocation of the finish task (the last level). After this allocation, the graph is traversed in a bottom-up fashion from level L − 1 to level 1. The mixed strategy starts by assigning the tasks belonging to the intermediate level, i.e. $k \in \{2, ...,L - 1\}$.

The mixed strategy starts by assigning the tasks belonging to the intermediate level, i.e. $k \in \{2, ...,L - 1\}$. Given starting level k, therefore the assignment of the tasks belonging to this level is only based on the execution cost

- **Pareto selection phase**

In this phase, we first compute the general completion time corresponding to each assignment and then only non- dominated solutions are maintained.

3.2 TIME BASED ALGORITHM

The time-based approach attempts to minimize the overall completion time (i.e. execution and communication time). As the cost-based approach, the time-based approach is an application matching and scheduling algorithm foran ”unbounded” number of VMs, which has three major phases, namely: a tasks sorting phase i), ii) an allocation phase and

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net

Vol-11 Issue-01 Sept 2016

iii) Pareto selection phase.

- **Tasks sorting phase**

This phase is the same as for the cost based algorithm. Recall that this phase allows to group the workflow application tasks that are independent of each other.

- **Resource allocation phase**

The cost-based and the time-based approaches differ mainly at resource allocation phase. Indeed, the first one approach focus on minimizing the cost function while the second approach attempts to minimize the time function. The objective of the resource allocation phase of the time-based approach is to choose the virtual machine that minimizes the finish time of each task and the task is assigned to that virtual machine. The three allocation strategies, detailed above, are also applied to explore this approach.

- **Pareto selection phase**

Recall that at the end of the previous phase L assignment of all tasks is obtained. In this phase, we first compute the overall completion time corresponding to each assignment and then only non-dominated solutions are maintained.

4. CONCLUSION

In this paper, we have proposed three bi-criteria complementary approaches to tackle the allocation and scheduling workflow problems in Cloud environments. Moreover, in order to assess the quality of obtained solutions by our approaches we have proposed two lower bounds for each considered criterion.

Unlike existing works, these approaches take into account two conflicting criteria simultaneously: i) execution cost and ii) execution time. Moreover, they offer more flexibility to consumer to estimate their preferences and choose a desired schedule from the obtained efficient solutions. More precisely, the first one focused on the cost incurred by using a set of resources, while the second approach attempts to minimize the overall execution time. The third approach is based on the two first approaches for selecting only the Pareto solutions.

Moreover, we plan to extend the proposed work to take into account others criteria like carbon emission and energy cost. In addition, it is interesting to adapt the proposed approaches to the case where a task cannot be executed on all virtual machine types.

REFERENCES

[1] The effective administration of data centers for cloud computing using an energy-

ijise.net

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net

Vol-11 Issue-01 Sept 2016

aware resource allocation heuristic was discussed in an article published in 2012 by R. Buyya, A. Beloglazov, and J. Abawajy in the journal Future Generation Computer Systems, pages 755–768. Software as a service, platform as a service, infrastructure as a service, and paaS are all mentioned in this article.

According to Wikipedia, data intensive computing is described in [3].

(Real Life data intensive applications: problems and solutions), 2010 Salishan Conference on high speed computing, J. Becla.

[5]The paper "Cloud brokering mechanisms for optimized placement of virtual machines across multiple providers" was published in 2012 by J. Tordsson, R. Montero, R. Moreno-Vozmediano, and I. Llorente in the journal Future Generation Computer Systems, volume 28, issue 2, pages 358-367.

The technical study "Policy-driven service placement optimization in federated clouds" was published in 2011 by IBM Haifa Labs and was written by D. Breitgand, A. Marashini, and J. Tordsson.

[7] J. Broeckhove, K. Vanmechelen, and R. V. den Bossche The most cost-effective scheduling approach for hybrid infrastructure as a service clouds for tasks that are time-sensitive, In Volume 3, Issue 5, Pages 228–235, 2010: Proceedings of the Third IEEE International Conference on Cloud Computing.

An article titled "Cost-Efficient Scheduling Heuristics for Deadline Constrained Workloads on Hybrid Clouds" was published in 2011 by R. Van den Bossche, K. Vanmechelen, and J. Broeckhove at the IEEE International Conference on Cloud Computing Technology and Science, with pages 320-327.

ijise.net