International Journal of Research

International Journal of Research (IJR)

e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 10, October 2015

Available at http://internationaljournalofresearch.org

Data Transformation in Cloud for Effective Cost Optimization

Chanda Anusha[1]& K.V.Rathnam [2]

[1]PG Scholar,Dept of CSE, Rao & Naidu Engineering College, Ongole,Prakasam Dist,Andhra Pradesh

[2]Assistant Professor,Dept of CSE, Rao & Naidu Engineering College, Ongole,Prakasam Dist,Andhra Pradesh

Abstract—

A general transformation-based optimization framework for workflows in the cloud. Specifically, ToF formulates six basic workflow transformation operations. An arbitrary performance and cost optimization process can be represented as a transformation plan (i.e., a sequence of basic transformation operations). All transformations form a huge optimization space. We further develop a cost model guided planner to efficiently find the optimized transformation for a predefined goal and our experimental results demonstrate the effectiveness of ToF in optimizing the performance and cost in comparison with other existing approaches. Cloud computing offers its customers an economical and convenient pay-as-you- go service model, known also as usage based pricing. Cloud customers pay only for the actual use of computing resources, storage, and band-width, according to their changing needs, utilizing the cloud's scalable and elastic computational capabilities. In particular, data transfer costs (i.e., bandwidth) is an important issue when trying to minimize costs.

Index Terms—Cloud computing; monetary cost Optimizations; workflows

I INTRODUCTION

Cloud Computing is a flexible, cost- effective, and proven delivery platform for providing business or consumer IT services over the Internet. However, cloud Computing presents an added level of risk because essential services are often outsourced to a third party, which makes it harder to maintain data security and privacy, support data and service availability, and demonstrate compliance. Cloud Computing has become more like an onset trend in the face of enterprises. It sets a milestone in the field of workflow execution in business process management. Workflow Management System is mainly devoted to support the definition as well as execution cum control of business processes. Workflow Scheduling is a key to workflow management. For efficient Scheduling in workflows, cost-time based evaluation of various algorithms has been included in this paper. First, users have different requirements on performance and cost. Some existing studies [5], [6] have focused on minimizing the cost while satisfying the performance requirement, some are aiming to optimize performance for a given budget [11] and others are considering the trade-off between performance and monetary cost [8], , [9]. Second, different cloud offerings result in significantly different cost structures for running the workflow. Even from the same cloud provider, there are multiple types of virtual machines (or instances) with different prices and computing capabilities. Third, workflows have very complicated structures and different computation/IO characteristics, as observed in the previous. We review the existing studies and find that most of them are *ad hoc* in the sense that they fail to capture the optimization opportunities in different user requirements, cloud offerings and workflows. For example, Kllapi et al. [8] consider only a single instance type. However, previous studies have shown that carefully selecting instance types is important forth overall cost. Mao et al. [5] focus on minimizing the cost while satisfying the performance requirement of individual workflows, and simply use a *fixed* sequence of workflow transformation operations. The fixed sequence can be effective for some of the workflows and cloud offerings, but ineffective for others. All those studies potentially lose optimization opportunities for performance and cost. We have identified three design principles. First, the framework should have an extensible design on the workflow optimizations, which adapts to different cloud offerings and workflow structures. Second, the framework should have a general optimization process for different requirements on performance and cost constraints. Last, the framework should be light weight for on-line decision making.

II PRELIMINARY AND RELATED WORK

In this section, we first describe cloud computing environments mainly from users' perspective.
Next, we present the security issues and review the related work

2.1 Cloud Environments

Cloud providers offer multiple types of instances with different capabilities such as CPU speed, RAM size, I/O speed and network bandwidth to satisfy different application demands. Different instance types are charged with different prices. Tables 1 and 2 show the prices and capabilities of four on-demand instance types offered by Amazon EC2 and Rack space, respectively. Amazon EC2 mainly charges according to the CPU, whereas Rack space mainly on the RAM size. Both cloud providers adopt the instance hour billing model, whereby partial instance hour usage is rounded up to one hour. Each instance has a non-ignorable instance acquisition time. Users usually have different requirements on performance and monetary cost [8].

2.2 Security Requirements

Confidentiality: outsourced data must be protected from the TTP, the CSP, and users that are not granted access. Integrity: outsourced data are required to remain intact on cloud servers. The data owner and authorized users must be enabled to recognize data corruption over the CSP side. Newness: receiving the most recent version of the outsourced data file is an imperative requirement of cloud-based storage systems. There must be a detection mechanism if the CSP ignores any data-update requests issued by the owner. Access control: only authorized users are allowed to access the outsourced data. Revoked users can read unmodified data, however, they must not be able to read updated/new blocks

2.3 Cost optimization for workflows

Performance and cost optimizations are a classic research topic for decades. Many scheduling and provisioning algorithms have been proposed leveraging market-based techniques, rule-based techniques [9] and models [7] for cost, performance and other optimization objectives.
Different applications re-quire different performance and cost optimizations. Many relevant performance and cost optimizations can be found in databases , Internet, distributed systems and so on. Performance and cost optimizations for workflows have been well studied on grid, cloud and heterogeneous computing environments. Specifically, we review the most relevant workflow optimizations in the grid and in the cloud. Performance and cost optimization techniques have been developed for a single cloud provider and multiple cloud providers. There have been much more research studies on a single cloud than on multiple clouds.

2.4 Application security

These applications are typically delivered via the Internet through a Web browser . However, flaws in web applications may create vulnerabilities for the SaaS applications. Attackers have been using the web to compromise user's computers and perform malicious activities such as steal sensitive data . Security challenges in SaaS applications are not different from any web application technology, but traditional security solutions do not effectively protect it from attacks, so new approaches are necessary . The Open Web Application Security Project (OWASP) has identified the ten most critical web applications security threats. There are more security issues, but it is a good start for

International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 10, October 2015
Available at http://internationaljournalofresearch.org

securing web applications. Infrastructure-as-a-service (IaaS) security issues

IaaS provides a pool of resources such as servers, storage, networks, and other computing resources in the form of virtualized systems, which are accessed through the Internet . Users are entitled to run any software with full control and management on the resources allocated to them . With IaaS, cloud users have better control over the security compared to the other models as long there is no security hole in the virtual machine monitor . They control the software running in their virtual machines, and they are responsible to configure security policies correctly. However, the underlying compute, network, and storage infrastructure is controlled by cloud providers. IaaS providers must undertake a substantial effort to secure their systems in order to minimize these threats that result from creation, communication, monitoring, modification, and mobility [42].

III Transformation Based Optimization

ToF has two major components for performance and cost optimizations: transformation model and planner. The trans-formation model defines the set of transformation operations for a workflow, and the planner performs The transformation on the workflow according to the cost model. Figure 1 shows the application model of ToF

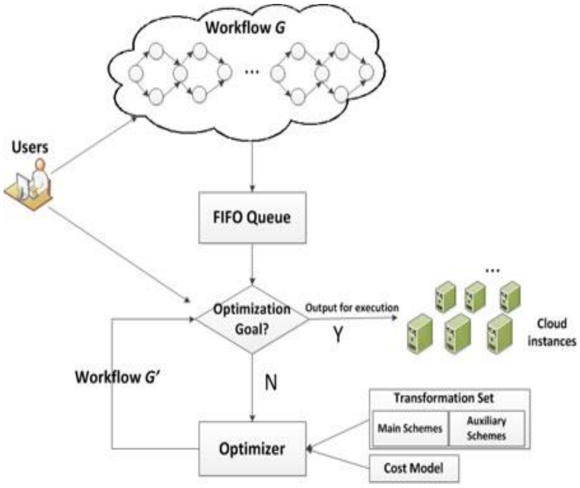

Fig. 1. System overview of ToF.

The planner is ran periodically. The period is denoted as the *plan period*. The workflows submitted during the epoch are temporarily stored in a queue, and then the planner performs optimization for all the workflows in the queue at the begin-ning of a plan period. This batch processing has two major benefits: first, maximizing the instance sharing and reuse so that the cost is reduced; second, reducing the planner overhead.

In each plan period, after all the workflows in the queue have been optimized, they are submitted to the cloud with their execution plan for execution. The execution plan is a set of instance requests. Each request includes the instance type, the starting and ending time
of the requested instance and the task(s) scheduled to the instance.

IV Optimization Sequence

a) ToF Planner

This ToF planner helps to find the sequence of operation to be performed for cost and performance optimization. The lanner has three designs. First the planner runs periodically. Second the main schemes and auxiliary schemes are performed alternatively and cost model avoids the unwanted transformation operations. Third the rule is defined to achieve the performance and cost optimization goal

b) The Optimization process of ToF for workflows

All workflows are allowed in the form of queue. Each workflow is initially assigned with instance type. The selected main scheme and auxiliary scheme operations are performed for largest cost reduction without affecting the deadline of the workflow. If the time constraints are not violated then the cost has been estimated by cost model. If these transformation operation has no cost reduction then the initial assignment are executed to optimize the workflow.

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 10, October 2015
Available at http://internationaljournalofresearch.org

c). Max-Min Task-Scheduling Algorithm

The main aim of task scheduling algorithm mainly in cloud computing is the allocation of task to the particular node. The allocation of task to the appropriate node is the challengeable task which can be improved by Max-Min Task Scheduling algorithm[6]. This scheduling algorithm maintains the task status table and virtual machine status table which help to identify the workload of virtual machines. This table also helps to find the task completion time.

When the tasks enter the scheduler, the scheduler finds the longest execution time tasks in the task status table then select the virtual machine in the virtual machine status table and finds the shortest completion time. This result in the allocation of tasks to the total number of tasks and their execution time in the virtual machine are updated in the virtual machine status table. This process is repeated until every task is allocated to the concerned virtual machine.

V Experimental setup

We conduct our evaluations on real cloud platforms, including Amazon EC2 and Rackspace

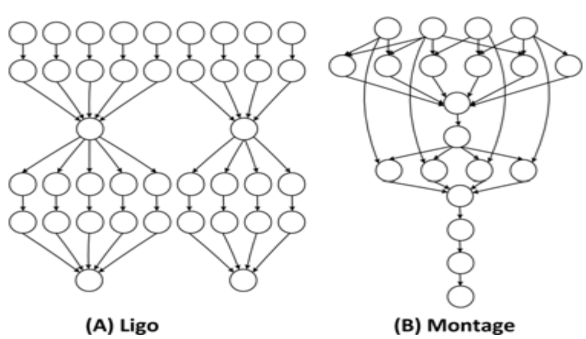

Fig. 2. Workflow structures of Ligo and Montage.

We have used synthetic workloads based on two realworld applications, namely Montage and Ligo. Montage is an astronomical application widely used as a Grid and parallel computing benchmark. Ligo (Laser Interferometer Gravitational Wave Observatory) is an application used to detect gravitational-wave. The two applications have different workflow Structures rate defined in Amdahl's law. The I/O and network time of workloads are included in the sequential part of the total execution time. In Section 6.3.2, we vary *Pr* to study the effectiveness of ToF on workloads with different I/O and network characteristics. The performance degradation rate for Co-scheduling operation is around 1:25 in our study. We define *Dmax* as the execution time of a workflow when its tasks are all assigned to the cheapest instance type while *Dmin* as the execution time of a workflow when all tasks are assigned to the most expensive instance type.

VI CONCLUSION

Performance and monetary cost optimizations for running workflows from different applications in the cloud have be-come a hot and important research topic. However, most existing studies fail to offer general optimizations to cap-ture optimization opportunities in different user requirements, cloud offerings and workflows. To bridge this gap, we pro-pose a workflow transformation-based optimization frame-work namely ToF. We formulate the performance and cost optimizations of workflows in the cloud as transformation and optimization. We further develop a cost model guided planner to efficiently and effectively find the suitable transformation sequence for the given performance and cost goal. We evaluate our framework using real-world scientific workflow applications and compare with other state-of-the-art scheduling algorithms. Results show our frame-work out performs the state-of-the-art Auto-scaling algorithm by 30% for monetary cost optimization, and by 21% for the execution time optimization. Moreover, the planner is lightweight for online optimization in the cloud environments. As for future work, we consider ToF on multiple clouds. Still, there are many practical and challenging issues for current multi-cloud environments . Those issues include relatively limited cross-cloud network bandwidth and lacking of cloud standards among cloud providers.

VII REFERENCES

[1]AmazonCaseStudies,http://aws.amazon.com/solutions/casestudies/,2014.

[2] Windows Azure Case Studies, "http//www.microsoft.com/azure/casestudies.mspx." 2014.

[3] S. Ibrahim, B. He, and H. Jin, "Towards Pay-As-You-Consume Cloud Computing," Proc. IEEE Int'l Conf. Services Computing (SCC), pp. 370-377, 2011.

[4] M. Mao and M. Humphrey, "Auto-Scaling to Minimize Cost and Meet Application Deadlines in Cloud Workflows," Proc. Int'l Conf. High Performance Computing, Networking, Storage and Analysis (SC), pp. 49:1-49:12, 2011.

[5] M. Mao, J. Li, and M. Humphrey, "Cloud Auto-Scaling with Deadline and Budget Constraints," Proc. IEEE/ACM 11th Int'l Conf. Grid Computing (GRID), pp. 41-48, 2010.

[6] E.-K. Byun, Y.-S. Kee, J.-S. Kim, and S. Maeng, "Cost Optimized Provisioning of Elastic Resources for Application Workflows," Future Generation Computer Systems, vol. 27, no. 8, pp. 1011-1026,
2011.

[7] H. Kllapi, E. Sitaridi, M.M. Tsangaris, and Y. Ioannidis, "Schedule Optimization for Data Processing Flows on the Cloud," Proc. ACM SIGMOD Int'l Conf. Management of Data, 2011.

[8] M. Malawski, G. Juve, E. Deelman, and J. Nabrzyski, "Cost- and Deadline-Constrained Provisioning for Scientific Workflow Ensembles in IaaS Clouds," Proc. Int'l Conf. for High Performance Computing, Networking, Storage and Analysis (SC), pp. 22:1-22:11, 2012.

[9] Y. Gong, B. He, and J. Zhong, "Network Performance Aware MPI Collective Communication Operations in the Cloud," IEEE Trans. Parallel and Distributed Syst., vol. 99, 2013, doi: 10.1109/TPDS. 2013.96.

[10] M. Mao and M. Humphrey, "Scaling and Scheduling to Maximize Application Performance within Budget Constraints in Cloud Workflows," Proc. IEEE 27th Int'l Symp. Parallel andDistributed Processing (IPDPS), 2013.

[11] H.M. Fard, R. Prodan, and T. Fahringer, "A Truthful Dynamic Workflow Scheduling Mechanism for Commercial Multicloud Environments," IEEE Trans. Parallel and Distributed Systems, vol. 24, no. 6, pp. 1203-1212, June 2013.