International Journal For Technological Research In Engineering
Volume 4, Issue 10, June-2017 *ISSN (Online): 2347 - 4718*

AUTONOMIC WORKFLOW MANAGEMENT IN CLOUD

Deeksha Satish[1], Prasanna B T[2]
[1]M.Tech 2nd Semester, [2]Assistant Professor
Department of Computer Science & Engineering, Sri Jayachamarajendra College of Engineering(SJCE),
JSS S&T University Campus, Mysore, Karnataka, India

Abstract: As many of the large-scale scientific applications executed on Grids are expressed as complex scientific workflows, workflow management has emerged as one of the most important Grid services in past few years. Scientific workflows can be defined as the aggregation of Grid application services, which are executed on distributed Grid resources in a well- defined order to satisfy the specific requirements of users. A workflow management system is generally employed to define, manage and execute these workflows in world-wide Grid environment. However, the increasing scale complexity, heterogeneity and dynamism of Grid environment that includes networks, resources and applications have made such workflow management systems brittle, unmanageable and insecure. Autonomic computing provides a holistic approach for the design and development of systems/applications that can adapt themselves to meet the requirements of performance, fault tolerance, reliability, security, etc., without manual intervention. Nowadays, workflows are the preferred means for the combination of services into added value service chains representing functional business processes or complex scientific experiments. A promising way to manage effectively services composition in a dynamic and heterogeneous environment is to make the workflow management framework able to self adapt at runtime to changes in its environment and provide an uniform resource access mechanism over Grid and Cloud infrastructures. Autonomic workflow management systems can support the runtime modification of workflows with the aim of improving their performance and recover from faults determining and provisioning the appropriate mix of Grid/Cloud services with requested QoS.

I. INTRODUCTION

Workflow management systems (WMSs) are generally utilizedto define, manage and execute workflow applications on Grid resources. WMSs must be dynamic and adaptive as Web/Grid services may be subjected to high variability in demand and suffer from unpredictable peaks of heavy load and/or the resources availability can change over time while the workflow is executing. When there are insufficient resources to meet the QoS requirement of the application, the system should provide additional computational resources. In these cases, the ability of temporarily extending the capacity of the Grid by renting resources from an external provider can be a viable proposition. Such an opportunity is offered by Cloud Computing which, by leveraging virtual machine technology, delivery IT infrastructure on demand on a pay peruse basis. Clouds allow users to acquire and release Services on-demand. These services can be Infrastructure as a Service(IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). Next generation computing environments will benefit from the combination of Grid and Cloud paradigms providing frameworks that integrate traditional Grid services with on demand Cloud services. This enablegrid workflow systems to easily grow and shrink the available resource pool as the needs of the workflow change over time. A workflow runningon a hybrid computing infrastructure should react to workload variations by altering its configuration in order to optimally use the Grids and Clouds available resources and recover from faults. Such modifications should happen automatically and without any human intervention. An autonomic WMSexhibits the ability to reconfigure itself to the changes in the environment and can adapt the size to keep the balance between servicing its workload with optimal performance and ensuring efficient resources allocation.

II. GRID COMPUTING

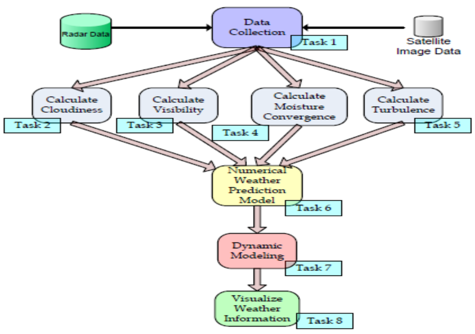
Over the last two decades, Grid computing [43] has emerged as one of the most promising technologies to build high performance distributed computing infrastructures. Computational Grids enable the sharing, selection, and aggregation of geographically distributed heterogeneous resources, such as computational clusters, supercomputers, storage devices, and scientific instruments. These resources are under control of different Grid organizations, offer huge computational power, and being utilized to solve many important scientific, engineering and business problems, such as protein folding, drug discovery, weather forecasting, earthquake engineering, financial modeling, and multi-player gaming.In general, Grid infrastructures are distributed, large, heterogeneous, uncertain, andhighly dynamic.Computational Grids can be categorized into different types, such as organizationGrids, enterprise Grids, and global Grids depending on the scope of resource sharing.global Grids provide means to connect the Internet-wide distributed Grid resources, administeredby multiple Grid sites with the aim of creating a collaborative computing environment.A sample scenario of resource sharing and application management in sucha global Grid environment is shown in Fig. 1.1, where application scientists and usersfrom different Grid sites/domains share various types of resources distributed over theworld. The applications executed in such environment are generally computation or dataintensive and composed of multiple tasks, where a task is a set of instructions that cae executed on a single processing element of a computing resource. Based on the dependencyor relationship among these tasks, Grid applications can be divided into threetypes: bag-of-task, tightly coupled message passing, and workflow applications. This paper presents a framework to facilitate efficient scheduling and management of workflow applications in global Grids.

III. WORKFLOW MANAGEMENT

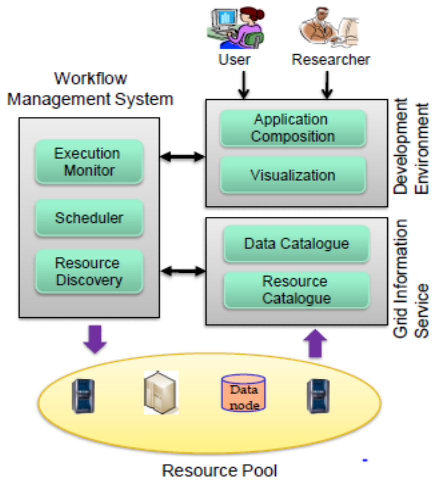
Many of the large-scale scientific applications executed on the present-day Grids are expressedas complex e-Science workflow [, which is a set of ordered tasks thatare linked by data dependencies (refer to Fig. 1.2(a)). A Workflow Management System(WMS) is generally employed to define, manage, and execute these workflow applications on Grid resources. Due to the continuous demand of leveraging high performance computing infrastructure for the execution of workflow applications by the scientist andresearch community, WMS has emerged as one of the most important Grid services inpast few years. Fig. 1.2(b) illustrates the execution of the workflow shown in Fig. 1.2(a)on a traditional distributed computing environment. A WMS uses a specific scheduling strategy for mapping the tasks in a workflow to suitable Grid resources in order to satisfy user requirements. Realizing a WMS in global Grids requires a number of challenges to be overcome, which include workflow applicationmodeling, resource discovery, task scheduling, information services, data management, and failure handling. However, the increasing scale complexity, heterogeneity, and dynamism of Grid environment have made such WMS brittle, unmanageable, and insecure. This thesis presents a framework for effectively managing workflow applications by devising a set of scheduling algorithms addressing these requirements.

IV. AUTONOMIC COMPUTING

Autonomic Computing (AC) is an emerging area of research for developing large-scale, self-managing, complex distributed system. The vision of AC is to apply the principles of self-regulation and complexity hiding for designing complex computer-based systems. Thus, AC provides a holistic approach for the development of systems that can adapt themselves to meet requirements of performance, fault tolerance, reliability, security, Quality of Service (QoS) etc. without manual intervention. systems. Thus, AC provides a holistic approach for the development of systems that can adapt themselves to meet requirements of performance, fault tolerance, reliability, security, Quality of Service (QoS) etc. without manual intervention.



(a) Example of workflow: weather prediction application



(b) Workflow management system

An autonomic workflow management system leverages the concept of AC and is ableto efficiently define, manage, and execute workflow applications in heterogeneous and dynamicGrid environment by continuously adapting itself to the current state of the system.Therefore, this thesis aims to design and develop algorithms and policies for building anautonomic workflow management system.

V. REPUTATION-BASED DEPENDABLE WORKFLOW SCHEDULING

A reputation-based Grid scheduling algorithm is presented to counter the effect of inherent unreliability and temporal characteristics of computing resources in large scale. The proposed algorithm can schedule workflows by dynamically adapting to changing resource conditions and offer significant performance gains as compared to traditional approaches in the event of unsuccessful job execution or resource failure. The results evaluated through an extensive trace driven simulation show that this scheduling technique can significantly reduce the make span and successfully isolate the failure-prone resources from the system.

VI. CONCLUSION

This thesis addresses the problem of enabling autonomic workflow management for Gridcomputing to effectively overcome the limitations of inherent uncertainty and dynamismof Grid environment. Autonomic policy based workflow management considers the abilityto self-reconfigure to the changes in the Grid environment; discover, diagnose andreact to the disruptions of workflow execution as well as monitor and optimize runtimeperformance automatically. By incorporating the self-properties of autonomic systemsinto the workflow management system, we demonstrated that it can effectively overcome the limitations of inherent uncertainty and dynamism of Grid environment as well as significantly improve the performance of workflow execution in global Grids. In this chapter, we first summarize the contributions and findings of this thesis, and then present a list of research issues for future investigations related to this work.

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, andM. Zaharia, A view of Cloud Computing, Communicationsof the ACM, 53(4):50-58, ACM, Apr. 2010.

[2] R. Buyya, C. S. Yeo, and S. Venugopal, Market-OrientedCloud Computing: Vision, Hype, and Reality for DeliveringIT Services as Computing Utilities, Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications (HPCC 2008), Sept. 25-27, 2008, Dalian, China.

[3] J. Varia, Best practices in architecting cloud applications inthe AWS Cloud. In: R. Buyya, J. Broberg, A. Goscinski(eds.), Cloud Computing: Principles and Paradigms, WileyPress, 2011.

[4] HP Utility Computing. http://www.hp.com/go/utility/.

[5] IBM Smart Cloud. http://www.ibm.com/cloud-computing/.

[6] J. O. Kephart and D. M. Chess, The Vision of Autonomic Computing, Computer, 36(1):41-50, IEEE, Jan. 2003.

[7] J. Kubiatowicz, D. Bindel, Y. Chen, S. Czerwinski, P.Eaton, D. Geels, R. Gummadi, S. Rhea, H. Weatherspoon,W. Weimer, C. Wells, and B. Zhao. OceanStore: anarchitecture for global-scale persistent storage, ACMSIGPLAN Notices, 35(11):190-201, ACM, Nov. 2000.

[8] J. Menon, D. A. Pease, R. Rees, L. Duyanovich, and B.Hillsberg. IBM Storage Tank—A heterogeneous scalableSAN file system, IBM Systems Journal, 42(2):250-267, IBM, Apr. 2003.

[9] K. Appleby, S. Fakhouri, L. Fong, G. Goldszmidt, M.Kalantar, S. Krishnakumar, D. P. Pazel, J. Pershing, and B.Rochwerger. Oceano-SLA based management of acomputing utility. Proceedings of the 2001 IEEE/IFIPInternational Symposium on Integrated NetworkManagement (INM 2001), Seattle, WA, USA, 2001.

[10] G. M. Lohman and S. S. Lightstone, SMART: making DB2(more) autonomic. Proceedings of the 28th InternationalConference on Very Large Data Bases (VLDB 2002), HongKong, China, 2002.

[11] H. Liu, V. Bhat, M. Parashar and S. Klasky, An Autonomic Service Architecture for Self-Managing Grid Applications, Proceedings of the 6th IEEE/ACM International Workshopon Grid Computing, Seattle, Washington, USA, 2005.

[12] M. Parashar and J.C. Browne, Conceptual and Implementation Models for the Grid, Proceedings of theIEEE, 93(3):653-668, IEEE, Mar. 2005.

[13] J. Kaufman, T. Lehman, G. Deen, J. Thomas, OptimalGrid -autonomic computing on the Grid, IBM, 2003

