ENABLING REPLICATION OF TASK ACROSS PUBLIC CLOUDS TO MEET DEADLINES OF SCIENTIFIC WORK FLOWS

[1] Mrs. D. Supriya, [2]Mrs. Greeshma, [3]Ms. V.Rohini
[1]Assistant Professor, [2] Assistant Professor, [3] Assistant Professor
[1] Department of ECE,
[1]P.S.R Engineering College, sivakasi , India

Abstract : This study has been undertaken to investigate the determinants of stock returns in Karachi Stock Exchange (KSE) using two assets pricing models the classical Capital Asset Pricing Model and Arbitrage Pricing Theory model. To test the CAPM market return is used and macroeconomic variables are used to test the APT. The macroeconomic variables include inflation, oil prices, interest rate and exchange rate. For the very purpose monthly time series data has been arranged from Jan 2010 to Dec 2014. The analytical framework contains**.**

IndexTerms - **: elasticity, cloud, EIPR, replication, budgets, workflow, speedup.**

I. INTRODUCTION

Cloud computing is a model for enabling convenient, on- demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. The Clouds layer provides support for modeling and simulation of cloud environments including dedicated management interfaces for memory, storage, bandwidth and VMs. It also provisions hosts to VMs, application execution management and dynamic system state monitoring. A cloud service provider can implement customized strategies at this layer to study the efficiency of different policies in VM provisioning. It provides elasticity to use pay-per-use system. Cloud environments do not present regular performance in terms of execution and also for data transfer times. It can cause performance variation of up to 30 percent for execution times and 65 percent for data transfer time. Scientific workflows is a specialized form of a workflow management system designed specifically to compose and execute a serious of computational or data manipulation steps or a workflow in a scientific application. Scientific workflows are described as direct acyclic graphs (DAGs) whose nodes represent tasks and vertices represent dependencies among tasks.

II METHODOLOGY

Our proposed algorithm is based on one of such algorithms (called IC-PCP), but also accounts for data transfer times and Cloud resources boot time during the provisioning and scheduling process. Furthermore, it explores possibility of tasks replication to increase the probability of meeting application deadlines. A scientific workflow application is modeled as a Direct Acyclic Graph (DAG) G= (T, ET) where T is the set of tasks that compose the workflow and E_T is the set of dependencies between tasks. Dependencies are in the form of edges $e_{i,j}=(t_i,t_j),t_i,t_j$£T that establish that a task t_j depends on data generated by t_i for its execution, and therefore t_j cannot start before execution of t_i completes and data generated by the latter is transferred to the location where t_j will execute. Task t_i is a parent task of t_j and t_j is a child task of t_i. Tasks without parents are called entry tasks and tasks without children are called exit tasks. A workflow can have only one entry task and one exit task. This can be achieved with the insertion of ''dummy'' tasks t_{entry} and t_{exit} that have execution time equals to 0. All the actual entry tasks are children of $t_{entry}$and all the actual exit tasks are parents of t_{exit}. The sets of parents and children of a task $t_j$are given respectively by functions parents (t_j) and children (t_j). Each workflow G has a soft deadline dl(G) associated to it. A Cloud provider offers a set of n virtual machine (VM) types denoted by= $vm_1,vm_2, ...,vm_n$. Each VM type offers different amount of resources, and incurs a different cost per use. Let C=c1, c2,….,cn be the cost vector associated with the use of each VM. VMs are charged per integer amount of time units, and partial utilization of a time period incurs charge for the whole period.

Deadline-constrained workflow scheduling algorithms for Infrastructure as a Service Clouds:

- In this paper use PCP algorithm for the Cloud environment and propose two workflow scheduling algorithms.
- Which aims to minimize the cost of workflow execution while meeting a user defined deadline.
- One-phase algorithm which is called IaaS Cloud Partial Critical Paths (IC-PCP)

Two-phase algorithm which is called IaaS Cloud Partial Critical Paths with Deadline Distribution (IC- PCPD).

III.FIRP ALGORITHM

The goal of the proposed Enhanced IC-PCP with Replication (EIPR) algorithm is increasing the likelihood of completing the execution of a scientific workflow application within a user-defined deadline in a public Cloud environment, which typically offers high availability but significant performance variation, with the use of task replication. In a high level, the proposed algorithm performs three distinct steps:

Step 1. Combined provisioned of Cloud resources and task scheduling

Step 2. Data transfer-aware provisioning adjust

Step 3. Task replication

IV .SYSTEM ARCHITECTURE

A system architecture as shown in fig 1 indicates the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.

A system architecture can consist of system components like deadline calculation, task selection, check deadline factor and the sub-systems like public cloud, data center developed, directed acrylic graph that will work together to implement the overall system. There have been efforts to formalize languages to describe system architecture, collectively these are called architecture description languages.

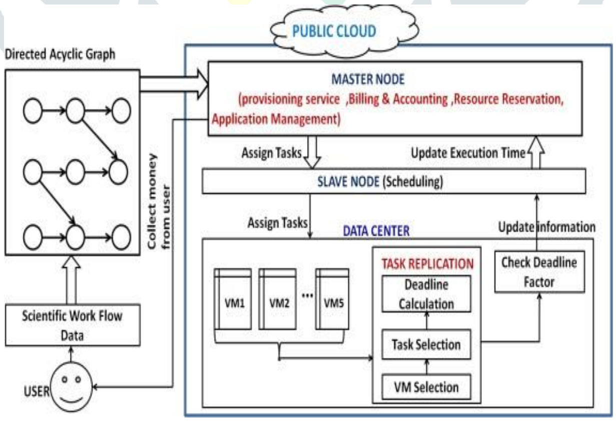

Fig 1 System architecture

V.MODULES DESCRIPTION

A. Cloudsim Setup Module

Clouds setup module provides support for modeling and simulation of cloud environments including dedicated management interfaces for memory, storage, bandwidth and VMs. It also provisions hosts to VMs, application execution management and dynamic system state monitoring.

B.Pre-processing modules

The main functionality of this module is to calculate all the parameters like metric, threshold, deadline of each task in a workflow etc., which are required in the process of scheduling workflows. The PM finds the threshold which is used as one of the parameter to prioritize the tasks in scheduling phase.

Threshold (Th) =Average no of children for the task in the workflow

C Replication based scheduling modules

This module sorts all tasks in the ready queue algorithm based on: Instructions_time_ratio is the ratio between the number of instructions in the task and the deadline of the task. The tasks with less instructions_time_ratio are scheduled first

D. Resubmission based executor

ResEM sends all mapped tasks to the respective data centers and also waits for the acceptance and reply from the data center. If the datacenter accepts the task, then ResEM sends the task to the data center by assigning a unique version identity to the task and then waits for the result.

V.SIMULATION RESULTS AND DISCUSSIONS

I A Start Simulation

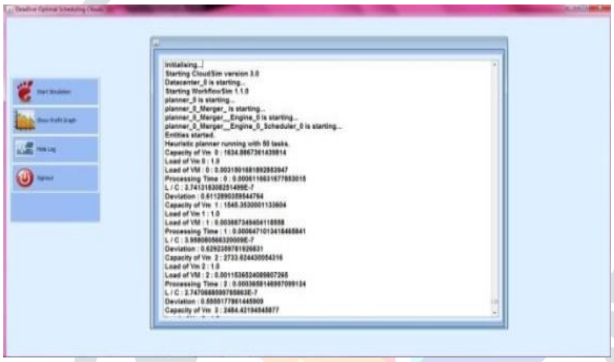

Fig 2 A.Start Simulation

The support for modeling and definition of cloud environment parameters like memory, bandwidth, vm's are indicated as in start simulation tool give in fig 2

2 Creating 5 VMs in Data Centre

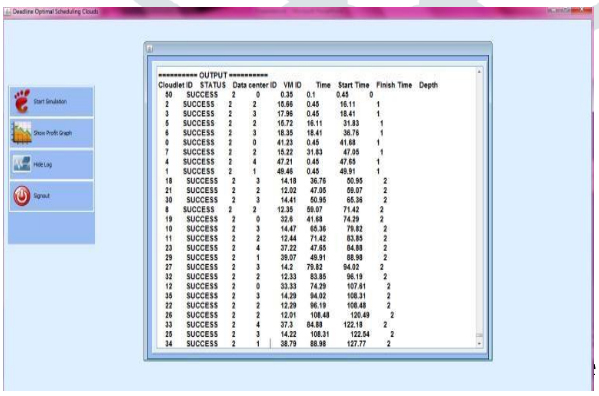

Fig 3 Creating 5 VM's in Data Centre

The 5 VM need to be set up in the data centre is to calculate all the parameters like metric, threshold, deadline of each task in a workflow etc., which are required in the process of scheduling workflows .Its simulation output is as shown in fig 3

3 Scheduling Result for 50 Tasks Using 5 VMs

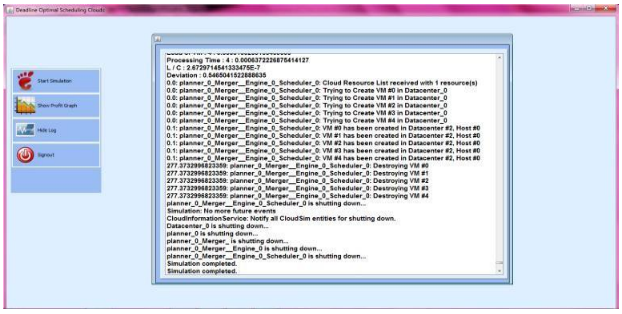

Fig 4 Result scheduling for 50 Tasks

The scheduling of tasks based on priority of the tasks to be executed like the instructions _time ratio number of services are done and waiting for the acceptance are to be simulated and the simulated result is as shown in fig 4

VI .CONCLUSION

Scientific workflows present a set of characteristics that make them suitable for execution in Cloud infrastructures, which offer on-demand scalability that allows resources to be increased and decreased to adapt to the demand of applications. However, public Clouds experience variance in actual performance delivered by resources. Thus, two resources with the same characteristics may have different performance in a given time, what results in variation in the execution time of tasks that may lead to delays in the workflow execution. To reduce the impact of performance variation of public Cloud resources in the deadlines of workflows, we proposed a new algorithm, called EIPR, which takes into consideration the behavior of Cloud resources during the scheduling process and also applies replication of tasks to increase the chance of meeting application deadlines. Experiments using four well-known scientific workflow applications showed that the EIPR algorithm increases the chance of deadlines being met and reduces the total execution time of workflows as the budget available for replication increases.

REFERENCES

1 Rodrigo N., "Meeting Deadlines Calheiros,RajkumarBuyya of Scientific Workflows in Public Clouds with Tasks Replication, '' IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 7, July 2014

2 G. Juve, A. Chervenak, E. Deelman, S. Bharathi, G. Mehta, and [2]K. Vahi, ''Characterizing and Profiling Scientific Workflows,'' Future Gener. Comput. Syst., vol. 29, no. 3, pp. 682-692, Mar. 2013

3 A.HiralesCarbajal, A.Tchernykh, R.Yahyapour, J.L.Gonzalez- Garcia, T. Ro ¨blitz, and J.M. Ramı´rez-Alcaraz, ''Multiple Workflow Scheduling Strategies with User Run Time Estimates on a Grid,'' J. Grid Comput., vol. 10, no. 2, pp. 325-346,June 2012

4 M. Mao and M. Humphrey, ''Auto-Scaling to Minimize Cost and Meet Application Deadlines in Cloud Workflows,'' in Proc. Int'l Conf. High Perform Computer. Netw., Storage Anal. (SC),2011,p.49

5 M. Rahman, X. Li, and H. Palit, ''Hybrid Heuristic for Scheduling Data Analytics Workflow Applications in Hybrid Cloud Environment,'' in Proc. IPDPSW,2011,pp.966-974

6 Z. Shi and J.J. Dongarra, ''Scheduling Workflow Applications on Processors with Different Capabilities,'' Future Gener.Comput. Syst., vol. 22, no. 6, pp. 665-675,May 2006

7 E.K.Byun, Y.-S. Kee, J.-S. Kim, and S. Maeng, ''Cost Optimized Provisioning of Elastic Resources for Application Workflows,'' Future Gener.Comput. Syst., vol. 27, no. 8, pp. 1011-1026,Oct 2011

8 W. Chen and E. Deelman ''WorkflowSim : A Toolkit for Simulating Scientific Workflows in Distributed Environments ''in Proc. 8th Int'l Conf. E-Science 2012, pp. 1- 8

9 S. Abrishami, M. Naghibzadeh, and D. Epema, ''Deadline- Constrained Workflow Scheduling Algorithms for IaaS Clouds,'' Future Gener.Comput. Syst., vol. 29, no. 1, pp. 158-169, , Jan. 2013

10 K. Plankensteiner and R. Prodan, ''Meeting Soft Deadlines in Scientific Workflows Using Resubmission.

11 Impact,'' IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 5, pp. 890-901, May 2012.