Workflow Management for Optimization the Scheduling in Grid Computing

Baljeet Kaur[1] Sarpreet Singh[2]
[1]Research Fellow [2]Asst. Professor
[1, 2] Department of Computer Science and Engineering
[1, 2]Sri Guru Granth Sahib World University, Fatehgarh sahib.

Abstract---Grid computing is one of the important computing techniques which are used for the storage of the large data. Grid computing which is a type of distributed computing is different from traditionally distributed system. There are various issues corresponding to the grid computing. The work flow is most critical issue in grid computing. The various resources which are available on the grid and using the technique of workflow these available are allocated to the uses. In the previous years, various techniques had been proposed for efficient work flow in grid computing. In this paper, we survey various techniques of work flow in grid computing.

Keywords: Work flow, Grid computing, cloud, cluster, and Fault tolerance.

I. INTRODUCTION

Computing is an activity which is used for creating and designing computers. Computing includes all the software and hardware designing, structuring and developing processes and manages them properly. Parallel and distributed are the types of computing.

In parallel computing, memory is shared between all the processors to exchange information. Bit-level, instruction level and data level are its type.

In distributed computing, each processor has its own private memory. Information is exchanged by passing messages between the processors. In this paper, survey is been provided of various techniques.

- Cloud computing
- Grid Computing
- Cluster computing

Cloud Computing is one of the types of distributed computing. Cloud computing is a branch of computing that comes under sharing of data instead of local servers. This policy of distributed computing is finished through pooling of all computer resources together. It is under the control of software not under any external users. Cloud computing is scalable process which runs over the virtual system with the help of service of internet. It is considered as a next generation of highly scalable distributed computing system. Cloud computing incorporates on-demand deployment, virtualization, open source software and Internet delivery of services. It also allowed users and its customers to access their personal files at any computer with the help of internet without installation.

Grid computing is hardware and software consisted infrastructure that gives a dependable, consistent and high capable computations at the end. it is different from cloud computing as it deals with homogenous system but grid computing is deal with heterogeneous type system which includes processors and workstations and clusters and gives different computational results. The resources are geographically distributed at different computer in grid computing. Grid computing allowing their computing power to be shared. Grids can merge the resources of thousands of different computers to make an especially powerful computing resource, so that it can be easily accessible from the comfort of a personal computer and can be easily useable. In grid every node has its own resource manager and works as a single entity.

Cluster computing is the third category that comes under distributed computing. Cluster computing is a form of computing in which a group of computers are linked together so that they can act like a single entity. Clusters are generally deployed to improve performance and availability over that provided by a single computer. It is usually much more cost-effective than single computers of comparable speed or availability. Components of all the clusters are mainly common [8]. As grid computing is heterogeneous in nature and it has no control over the locally available resource because its resources are distributed geographically. So information becomes limited due to no control over local resources. Current scheduling system in grids is time-dependent and cannot meet the necessary requirements like network allocation etc. This is necessary for the better performance of the system.

II. WORKFLOW SCHEDULING IN GRID COMPUTING

Grid computing is one of the important computing techniques which are used for the storage of the large data. Grid computing which is a type of distributed computing is different from traditionally distributed system**.** Security, resource allocation and scheduling are the major challenges in grid computing [9]. In grid computing scheduling is not as simple as it is in multiprocessor systems due to some factors like resource allocation and customer's interests etc. Grid computing is geographically distributed networks so in remote distributed network scheduling is tough job. Workflow scheduling is one of the types of scheduling in grid computing [10]. A workflow scheduler should have two capabilities one is resource allocation which distributes tasks to multiple resources and second is task execution and coordination which submits tasks to the resource's local schedulers in the right order and handles task dependencies so that no deadlock state occurs. A workflow scheduler in a grid is also a grid scheduler that works on top of multiple local schedulers. The grid scheduling process makes resource allocation decisions involving resources over multiple administrative domains, which introduce several challenges for workflow scheduling. The task dependencies find out the order of task submission and file transfer, which is the topological order of the workflow directed acyclic graph (DAG). In this order, initial start-time of each task can be calculated easily, as long as we know when the workflow

itself should be started. An allocated resource for a task should be available before its earliest start-time so that no delay is incurred because of the unavailability of resources.

III. Literature Review

Hassan [1] proposed workflow management based on a multi-objective Genetic Algorithm (GA) to improve grid computing performance. Task runtime is an important factor in grid computing. The proposed method considers a collection of levels as workflow so that to reduce the need to check workflow dependencies after a solution is obtained for the next population. As a result of this both time of scheduling and solution quality are improved. Results are presented which show that the proposed method has better performance compared to similar techniques.

Sandip et al. [2] proposed an Enhanced Genetic Algorithm (EGA) for achieving task scheduling with load balancing. The Enhanced Genetic Algorithm (EGA) is designed based on the standard Genetic Algorithm (GA).The method requires an encoding scheme which can represent all legal solutions to the optimization problem. The simulation results show that proposed algorithm yields better performance when compared with other traditional heuristic approaches.

Reza et.al [3] proposed to assign the tasks to the grid resources with goal of minimizing the total make span of the environment. The algorithm uses genetic approach to find the most suitable match between the tasks and grid resources. The simulation results obtained after applying the proposed algorithm is to schedule independent and sequential tasks to the grid resources so that it displays the applicability of the algorithm in grid environments.

Pooranian [4] et al. combine the genetic algorithm and Gravitational Emulation Local Search (GELS) as a method to solve scheduling problem by which simultaneously pay attention to two factors of time and number of missed tasks. Results shows that the proposed algorithm can decrease make span while minimizing the number of missed tasks compared with the traditional methods. The purpose of Grid computing is to utilize computational power of idle resources which are distributed in different areas. Since task scheduling includes in the NP-hard problems various researches have focused on invented algorithms especially the genetic ones. But since genetic is an inherent algorithm which searches the problem space globally and does not have the efficiency required for local searching.

Amin et.al [5] proposed approach gravitational attraction search as a local search algorithm has been associated to genetic algorithm to enhance its capability to search more intelligent in problem search space and achieve accurate response in less time. Comparing Hybrid Genetic Algorithm-Gravitational Attraction Search algorithm (HYGAGA) and genetic algorithm results asserts significant enhancement in the performance of search algorithm. In addition, Hybrid Genetic Algorithm-Gravitational Attraction Search algorithm (HYGAGA) could attain more appropriate solutions with comparing to other genetic algorithms like Genetic Algorithm (GA) and Hill Climbing (HC), Genetic Algorithm (GA) and Simulated Annealing (SA) and Genetic Algorithm (GA) and Tabu Search (TS) implemented and analyzed to compare their ability in search in problem search spaces. The Hybrid Genetic Algorithm-Gravitational Attraction Search, HYGAGA, proposed as a memetic algorithm to solve grid task scheduling problem.

IV. Problem Formulation

Task scheduling and resource allocation are two of the most important issues in grid computing. In a grid computing system, the workflow management system receives inter-dependent tasks from users and allocates each task to an appropriate resource. The assignment is based on user constraints such as budget and deadline. Thus, the workflow management system has a significant effect on system performance and efficient resource use. In earlier work genetic algorithm or ant colony optimization used to get the optimal solution. So to remove this problem new hybrid approach is applied. This hybrid approach is the combination of genetic algorithm and ant colony optimization which will gives better workflow management than previous one. So efficiency and performance of the system is also improved than previous one. It gives better quality workflow scheduling.

V. Conclusion And Future Work

Grid computing is a computing environment with high performance to solve larger scale computational demands. Allocation of resources, security, scheduling are the key challenges of grid computing. The two approach genetic algorithm and ant colony optimization has been implemented to obtain the optimal solution. But there is a problem of workflow scheduling which is not achieved by using these above said techniques and gives poor performance of the system. The various disadvantages of traditional approach have been discussed in this paper. In future, we will implement Hybrid technique which is composition of genetic and ant colony optimization for better performance so that workflow management system is more fast and efficient than traditional approaches.

References

[1] Hassan Khajemohammadi, Ali Fanian, T. Aaron Gulliver, "Fast Workflow Scheduling for Grid Computing Based on a Multi-objective Genetic Algorithm" IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM), pp no. 96 – 101, 2013.

[2] Dipti Sharma Mr. Pradeep Mittal, "Job Scheduling Algorithm for Computational Grid in Grid Computing Environment" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 5, May 2013.

[3] Reza Entezari-Maleki, Ali Movaghar, "A Genetic Algorithm to Increase the Throughput of the Computational Grids" International Journal of Grid and Distributed Computing, Vol. 4, No. 2, June, 2011.

[4] SaeedFarzi, "Efficient Job Scheduling in Grid Computing with Modified Artificial Fish Swarm Algorithm" International Journal of Computer Theory and Engineering, Vol. 1, No. 1, April 2009, 1793-8201.

[5] Z. Pooranian, A. Harounabadi, M. Shojafar and N. Hedayat, "New Hybrid Algorithm for Task Scheduling in Grid Computing to Decrease missed Task" World

Academy of Science, Engineering and Technology 55 2011.
[6] Amin. Jula, NarjesKhatoon. Naseri, "A Hybrid Genetic Algorithm-Gravitational Attraction Search algorithm (HYGAGA) to Solve Grid Task Scheduling Problem" International Conference on Soft Computing and its Applications (ICSCA'2012) August 25-26, 2012 Kuala Lumpur (Malaysia).
[7] Andrew S. Tanenbaum and Maarten Van Steen, Distributed Systems: Principles and Paradigms, Pearson Prentice Hall, 2nd Edition 2007.
[8] Cong Wang "Towards Secure and Dependable Storage Services in Cloud Computing" IEEE, 2012
[9] Raksha Sharma, Vishnu Kant Soni, Manoj Kumar Mishra, PrachetBhuyan, "A Survey of Job Scheduling and Resource Management in Grid Computing" World Academy of Science, Engineering and Technology 40 2010.
[10] Phillipa Sessini, "Scheduling in Grid Computing Systems" CPSC 531 Term Project, August 2005.

