RESEARCH ARTICLE OPEN ACCESS

A Survey on Scheduling Strategies in Cloud Computing

Aarti [1], Malti Rani [2], Lohit Kapoor [3]
Punjab Institute of Technology [1] & [2]
Kapurthala (PTU MAIN CAMPUS)
Punjab
Thaper University [3], Patiala
India

ABSTRACT

Cloud computing is an emerging model and it allows customer to pay as you demand and has the huge achievement. Cloud computing is a heterogeneous system as well and it holds large amount of operation data. In the scheduling mechanism some demanding data or computing demanding application, it is approved that optimizing the transmitting and processing time is vital to an application program. By virtue of comparing various algorithm in crossover and mutation and in the local research, and their experiment results. In third paper we put across various issues in cloud computing and identified the issues that needs to be immediately caters.

***Keywords*:-** Scheduling, Optimization, Cloud.

I. INTRODUCTION

The modern remote sensing quantitative retrieval is a massive data processing process, the need for a large number of data preprocessing and the calculation for complex radiation transfer equations, which is time-consuming. The use of ordinary computer needs a few days for it. This situation has seriously affected the large number of practical applications of remote sensing data. Cloud computing is an emerging technology that allows users to utilize on-demand computation, storage, data and services from around the world. Workflow processing services based on cloud computing platforms can access to a wide variety of computing resources from a distributed environment by the network, and provide services by scheduling different geographic locations of remote sensing algorithms and remote sensing data in accordance with the processing rules.

The SOA architecture of cloud computing and its on-demand computing service can be very good to meet the requirements of the mass remote sensing data storage and computation, thus it is necessary to study the cloud-based computing platform for the processing services of remote sensing quantitative retrieval workflow. An IaaS cloud data centre is used to host computer systems and all the related components. Cloud computing data centres are emerging as new applicant for restore traditional data centres that are growing rapidly in both number and capability to convene the increasing load for computing resources and storages. Rapid enlargement of the request for computational power by scientific, business and web-applications has pave the way to the establishment of large-scale data centers. The research of quantitative remote sensing workflow processing services based on Cloud computing platform is rarely introduced.

There are some related literatures, which introduced static directed acyclic graphs (DAGs) for grid workflow scheduling. And many heuristic scheduling algorithms for heterogeneous environment have been proposed. Because most of these hypothetical calculations and cost of communication have precise forecast, this limits the above algorithms in dynamic environment applicability. Therefore, in order to improve the performance of the application, the initial scheduling strategy should be based on the assessment, but the running (run-time) should be rescheduling [Altintas, et al., 2004]. In order to solve the recursive problems in a lot of scientific computing, Prodan et al. [2005] proposed a workflow mixing scheduling method based on directed graph.

The method is also used in computing and grid resource dynamic environment. Their dynamic scheduling algorithm is based on the iteration for the classical static DAG scheduling algorithm. In order to locate resources, a new search method is proposed by Wu [Wu, et al., 2005]. The method can efficiently find the optimal grid workflow scheduling scheme. The authors of algorithms used data acquisition and task operation to the modeling of workflow execution, and proposes a optimization scheduling algorithm based on simulated annealing. Sakellariou and Zhao [2004] proposed a low cost new scheduling strategy. The scheduling algorithm can effectively reduce the cost, and the algorithm be realized by rescheduling some chosen point in the implementation

II. CHALLENGES IN SCHEDULING

The key challenges that have addressed are:

2.1) How to optimally solve the trade-off between energy savings and delivered performance?

2.2) How to determine when, which VMs, and where to migrate in order to minimize energy consumption by the system, while minimizing migration overhead and ensuring SLA?

2.3) How to develop efficient decentralized and scalable algorithms for resource allocation?

2.4) How to develop comprehensive solution by combining several allocation policies with different objectives?

When these algorithms are applied to the field of remote sensing, they are difficult to meet the massive data character of remote sensing. Massive data processing requires us to minimize communication transmission. For example, we process a day data of the aerosol quantitative retrieval calculation, which need to deal with 13GB data. Using the conventional workflow scheduling algorithm in the processing of massive remote sensing data, task execution engine and task submitted machine kept transmitting data. During this process, the transmitted data include control data, the remote sensing original data, intermediate processing results, and the final results of multiple processing steps. The massive data were repeated transmission among machines. When we use this strategy to process a day data of the aerosol quantitative retrieval calculation, the transmission of data will be of a GB data transmission unit in the network. It is difficult for us to obtain quick processing performance. So we need urgently to solve a problem which is to reduce the amount of data processing of remote sensing in the network transmission.

III. GENERAL TRENDS

Massive data scheduling processing design in the cloud platform are as follows:

3.1) remote sensing data and processing algorithms are distributed to store using the oracle database in the cloud environment, and customized workflow model by user is sent to a computer with the highest performance in the cloud environment;

3.2) Selected high-performance computers obtain the corresponding remote sensing data, algorithm and the description of algorithm from the oracle database, in accordance with remote sensing process data and algorithm defined by customized workflow model；

3.3) Workflow scheduling engine obtain the description information of algorithm from the oracle database, and analysis of remote sensing algorithm to determine the time of calculation in a single high-performance computer. Then remote sensing data is divided according to the independent of each other. Workflow scheduling engine access the compute nodes from the control center, which can meet the operating conditions. Using the algorithm of network transmission to test the speed of p2p, depending on the size of the amount of data to be processed and network speed tested, it can estimate the time of the network transmission, the network transmission. If the transmission time is less than the computation time of the processing steps, Workflow scheduling engine select one of the worst compute nodes, which can meet the operating conditions. Workflow scheduling engine calculate the time of the maximum amount of data processed by the compute nodes, and estimate the time of calculation result to be returned；

3.4）If the calculation time on different computers to calculate is more than the calculation time on a single high performance computers to calculate, Workflow scheduling engine will select high-performance computers to calculate.

3.5）If the calculation time on different computers to calculate is less than the calculation time on single high-performance computers to calculate, the scheduling management center of the cloud controller will schedule the step in parallel on cloud environment.

In figure 1 we can see a sample working of cloud service provider in which the requests are allocating dynamically and the workload is handled by the users.

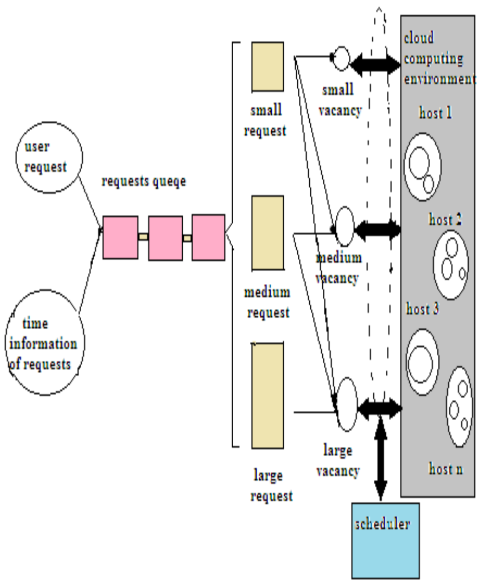

Figure 1: System Framework

Obviously, the allocation must meet the following constraints:

(1) That each virtual machine instance needs to be assigned and can only be assigned to one matching virtual machine instance vacancy.

(2) If two virtual machine instances using the same hardware resources, there should have some time interval between two usages. Because after the running of a virtual machine is completed, time is needed to process some operations like data clean and virtual machine restart.

(3) Virtual machine instance needs and the type of virtual machine instance vacancies should match each other.

(4) When large or medium vacancy is filled by small instances, we can release some new vacancies according to the proportion. For example, after a small instance fills a large vacancy, a medium-sized vacancy and two small vacancies can be produced. On the basis of above mentioned points we compared few algorithm we compare few algorithm and their results:

Task scheduling is very important to scientific workflows and task scheduling is challenge problems too. It has been research before in traditional distributed computing systems. Reference [9] is a scheduler in the Grid that guarantees that task scheduling activities can be queued, scheduled, monitored and managed in a fault tolerant manner. Reference [10] proposed a task scheduling strategy for urgent computing environments to guarantee the data's robustness. Reference [11] proposed an energy-aware strategy for task scheduling in RAID structured storage systems. Reference [12] studies multicore computational accelerators and the MapReduce programming model for high performance computing at scale in cloud computing.

They evaluated system design alternatives and capabilities aware task scheduling for large-scale data processing on accelerator-based distributed systems. They enhanced the MapReduce programming model with runtime support for utilizing multiple types of computational accelerators via runtime workload adaptation and for adaptively mapping MapReduce workloads to accelerators in virtualized execution environments. However, none of them focuses on reducing the processing cost and transmitting time between data centers on the Internet. As cloud computing has become more and more important, new data management systems have designed, such as Google's GFS (Google File System) and Hadoop. Their data hide in the infrastructures and the users can not control them. The GFS is designed mainly for Web search applications. Some researchers are based on cloud computing.

The Cumulus project [13] introduced scientific cloud architecture for a data centre. And the Nimbus [14] toolkit can directly turn a cluster into a cloud and it has already been used to build a cloud for scientific applications. Within a small cluster, data movement is not a big problem, because there are fast connections between nodes, i.e. the Ethernet and the processing time is not longer. However, the scientific cloud workflow system is distributed applications which need to be executed across several data centers on the internet. In recent studies, Reference [15] from the cost aspect studied the compute-intensive and data-intensive application. They formulate a non-liner programming model to minimize the data retrieval and executing cost of data-intensive workflows in clouds. Reference [16] investigated the effectiveness of rescheduling using cloud resources to increase the reliability of job completion. Specifically, schedules are initially generated using grid resources while cloud resources are used only for rescheduling to deal with delays in job completion. A job in their study refers to a bag-of-tasks application that consists of a large number of independent tasks; this job model is common in many science and engineering applications. They have devised a novel rescheduling technique, called rescheduling using clouds for reliable completion and applied it to three well-known existing heuristics. Reference [17] proposed matrix based k-means clustering strategy to reduce the data movement in cloud computing.

However, the reducing of data movement and cost do not mean that the processing cost and transmitting time decrease. In this work, we try to schedule the application data based on PSO algorithm in order to reduce the data transmitting time and process cost. Reference [18] study the deployment selection challenge from two different and usually conflicting angles, namely from the user's and the system provider's perspective. Users want to optimize the execution of their specific requests without worrying about the consequences for the overall system. The provider's objective however is to optimize the system throughput and allow a fair usage of the resources, or a usage mode as defined by the decision makers. While the users are most likely pursuing the same strategy for each request, the system responsible may face a dynamic environment, including changing requirements, changing usage patterns and changing decisions in terms of business objectives.

To address this issue, they propose a multi objective optimization framework for selecting distributed deployments in a heterogeneous environment based on Genetic Algorithm (GA). In fact, task assignment has been found to be NP complete [19]. Since task assignment is NP-Complete problem, Genetic Algorithm (GA) has been used for task assignment [20]. But, genetic algorithm may not be the best method. Reference[21] has illustrated that the particle swarm optimization algorithm is able to get the better schedule than genetic algorithm in grid computing. Reference [22] has shown that the performance of Particle Swarm Optimization (PSO) algorithm is better than GA algorithm in distributed system.

IV. CONCLUSION

In many different domains, in order to improve the efficiency the optimizing task scheduling is necessary. In cloud computing resources distribute all over the world, and the data usually is bigger and the bandwidth often is narrower, these problems are more important. In this paper, we presented a comparison in the task scheduling optimizing method in cloud computing. By comparing and analyzing mutation and local search algorithm based on particle swarm, which represents better performance.

In future work, our research is to center on the energy efficiency and service availability in cloud computing system. We aim to improve task scheduling optimization and make optimization policy to optimize not only the efficiency, but also the energy and service level agreement.

REFERENCES

[1] E. Deelman, J. Blythe, Y. Gil, C. Kesselman, G. Mehta, S. Patil, M.-H. Su, K. Vahi,M. Livny, Pegasus: mapping scientific workflows onto the grid, in: European Across Grids Conference, Nicosia, Cyprus, 2004, pp. 11–20.

[2] B. Ludascher, I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E.A. Lee,Scientific workflow management and the Kepler system, Concurrency and Computation: Practice and Experience (2005) 1039–1065.

[3] T. Oinn, M. Addis, J. Ferris, D. Marvin, M. Senger, M.Greenwood, T. Carver, K.Glover, M.R. Pocock, A. Wipat, P.Li, Taverna: a tool for the composition and enactment of bioinformatics workflows, Bioinformatics 20 (2004) 3045–3054.

[4] E. Deelman, A. Chervenak, Data management challenges of data-intensive scientific workflows, in: IEEE International Symposium on Cluster Computing and the Grid, (2008) 687–692.

[5] B. Ludascher, I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E.A. Lee, Scientific workflow management and the Kepler system, Concurrency and Computation: Practice and Experience (2005) 1039–1065.

[6] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, andI. Brandic. Cloud computing and emerging it platforms: Vision,hype, and reality for delivering computing as the 5thutility. Future Generation Computer Systems,(2009) 25(6):599–616 .

[7] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. H. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. Above the Clouds: A Berkeley View of Cloud Computing http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html

[8] I. Foster, Z. Yong, I. Raicu, S. Lu, Cloud computing and grid computing 360-degree compared, in: Grid Computing Environments Workshop, GCE'08, (2008) 1–10.

[9] T. Kosar, M. Livny, Stork: Making data placement a first class citizen in the grid, in: Proceedings of 24^{th} International Conference on Distributed Computing Systems, ICDCS 2004,(2004) 342–349.

[10] J.M. Cope, N. Trebon, H.M. Tufo, P. Beckman, Robust data placement in urgent computing environments, in: IEEE International Symposium on Parallel &Distributed Processing, IPDPS 2009, (2009)1–13.

[11] T. Xie, SEA: A striping-based energy-aware strategy for data placement in RAID-structured storage systems, IEEE Transactions on Computers 57 (2008) 748–761.

[12] M. Mustafa Rafique a, ∗ , Ali R. Butt a, Dimitrios S. Nikolopoulos b,c, A capabilities-aware framework for using computational accelerators in data-intensive computing, J. Parallel Distrib. Comput. 71 (2011) 185–197

[13] L. Wang, J. Tao, M. Kunze, A.C. Castellanos, D. Kramer, W. Karl, Scientific cloud computing: Early definition and experience, in: 10th IEEE International Conference on High Performance Computing and Communications, HPCC'08,(2008) 825–830.

[14] K. Keahey, R. Figueiredo, J. Fortes, T. Freeman, M.Tsugawa, Science clouds:Early experiences in cloud computing for scientific applications, in: First Workshop on Cloud Computing and its Applications, CCA'08,(2008) 1–6.

[15] S. Pandey, A. Barker, K. K. Gupta, R. Buyya ,Minimizing Execution costs when using globally distributed cloud services,2010 24th IEEE International Conference on Advanced Information Networking and Applications

[16] Y. C. Lee, A. Y. Zomaya, Rescheduling for reliable job completion with the support of clouds, Future Generation Computer Systems 26 (2010) 1192_1199

[17] D Yuan, Y Yang, X Liu, A data placement strategy in scientific cloud workflows, Future Generation Computer Systems(2010)1200-1214

[18] E. Vineka, P. P. Beranb, E. Schikutab, A dynamic multiobjective optimization framework for selecting distributed deployments in a heterogeneous environment , Procedia Computer Science 4 (2011) 166–175

[19] V.M. Lo, "Task assignment in distributed systems", PhD dissertation, Dep. Comput. Sci., Univ. Illinois, Oct. 1983.

[20] G. Gharooni-fard, F. Moein-darbari, H. Deldari and A.Morvaridi, Procedia Computer Science, Volume 1, Issue 1, May 2010,Pages1445-1454 ,ICCS 2010.

[21] L. Zhang, Y.H. Chen, R.Y Sun, S. Jing, B. Yang. " A task scehduling algorithm based on PSO fro Grid Computing", International Jouranal of Computational Intelligence Research.(2008),pp.37-43.

[22] A. Salman. "Particle swarm optimization for task assignment Problem". Microprocessors and Microsystems, November 2002. 26(8):363–371.

[23] Kennedy J, Eberhart RC (1995), Particle swarm optimization. In:Proceedings IEEE

Int'l. Conf. on Neural Networks, vol. IV,1942– 1948.

[24] M. Fatih Tasgetiren, Yun-Chia Liang, Mehmet Sevkli, and Gunes Gencyilmaz, "Particle swarm optimization and differential evolution for single machine total weighted tardiness problem," International Journal of Production Research, (2006) 4737-4754

[25] Y Shi, R C Eberhart. Empirical study of particle swarm optimization. Proc. IEEE Congr. Evol. Compute. (1999)1945-1950.