ENHANCEMENT OF DATA ORIENTED GRID SCHEDULING USING DYANAMIC FAULT TOLERANCE

Radha B.[1], **Sumathy V.**[2]

[1]Department of Computer Applications, Sri Ramakrishna Engineering College, Coimbatore, India
[2]Department of Electronics and Communication, Government College of Technology, Coimbatore, India.
Email: [1]radhakbr10@gmail.com

ABSTRACT

Traditional distributed computing systems closely couple data handling and computation The key features of the first batch scheduler specialized in data placement and data movement is Stork. Stork is especially designed to understand the semantics and characteristics of data placement tasks, which can include data transfer, storage allocation and de-allocation, data removal, metadata registration and replica location. The Stork also has its own drawbacks in detecting the failures resulting from back-end system level problems, like connectivity failure which is technically untraceable by users. Error messages are not logged efficiently, and sometimes are not relevant/useful from users' point-of-view. Our study explores the possibility of efficient error detection and reporting system for such environments. Besides, early error detection and error classification have great importance in organizing data placement jobs. It is necessary to have well defined error detection and error reporting methods to increase the usability and serviceability of existing data transfer protocols and data management systems.

Keywords Distributed systems, data aware scheduling, Error Detection, Grid computing, performance of systems, Scheduling.

I. INTRODUCTION

The latency and the throughput are the two main factors for performance in the closely coupled distributed environment. Failure during the data transfer is very common for example the user may not be aware of the background network connectivity failure[4]. The importance of error propagation and categorization of errors in Grid computing has been mentioned clearly in [5]The users of the system or the distributed environment may not be aware of what has been went wrong during their data transfer .This paper focuses on how to transfer data efficiently with out any disturbance in transfer. The errors are detected prior to data transfer and an alternative service to data transfer is suggested. Here we focus on what sort of an error has occurred and whether the destination node can be reached for transfer of data and how to classify those errors based on their classification ,and the performance of the system with and without prior error detection. There has been many efforts to implement file transfer protocols over distributed environments conforming to the security framework of the overall system. These solutions should ideally exploit communication channel to tune-up network and to satisfy high throughput and minimum transfer time [8,9]. Parallel data transfers, concurrent connections, and tuning network protocols such as setting TCP buffer are some of the techniques applied [7].The distributed environment differs from the normal network environment by its network topological structure. The band width and the latency are the two main characteristics which affect the performance of the data transfer in the distributed environment.

II. DATA SCHDEULING

There are many algorithms like genetic algorithms and heuristic techniques which are available to schedule data. In the previous work [1] of my research a comparison was made on those algorithms for their performance in the data scheduling and the best optimal algorithm was chosen .Even those algorithms work fine they have their own drawbacks which was rectified by using the techniques like Stork[2] which is specialised in data placement and data movement. This scheduler uses the job description language for data placement jobs .It also can interact with high level planners and workflow mangers in order to verify when and in which network to send to send the data. The Stork has also it s drawbacks like congestion in the network through which the data is transmitted to the destination node[6] .This sort of problem is rectified using our methodology. An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

The work flow model of the data and grid resource scheduling is given in the figure1 clearly [3]

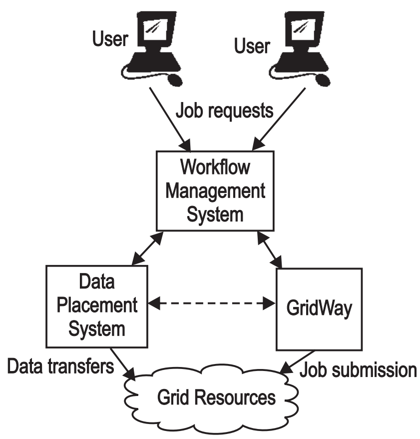


Fig. 1. Workflow management system to schedule job and data

As per the diagram the users job request is processed by the workflow management system which manages the job submission to the grid resource through the Gridway and the transfer of data through the data placement system[10].

III. FAILURE DETECTION

When data is been transmitted through a network there may occur an error while transmission of data and an message statement may be sent to the sender stating the actual problem that happened while transmission. The failures will be like the remote host server may be down, or file transfer service is not functioning in the host, or file transfer service is not supporting some of the features requested, there may be a mal-functionality in the service protocol, or user credentials are not satisfied, or any other problem occurred in the source server. In addition it is also necessary to verify whether destination host and the service is available or not so that data transfer to that particular destination will not be processed until the errors are rectified. As well the information about the active services in the node will help to choose the alternative protocols for transfer.

IV. EVALUATION AND DISCUSSION

In this work we proceed in such a way to prove that during the data transfer mechanism the data is transferred to the destination with prior error detection works effectively compared to the normal data transfer with the scheduler tools available.

A. Data Aware Scheduler

In this work first we have worked with the data transfer mechanism and with the code we developed the scheduler activity which is shown in Fig 2.

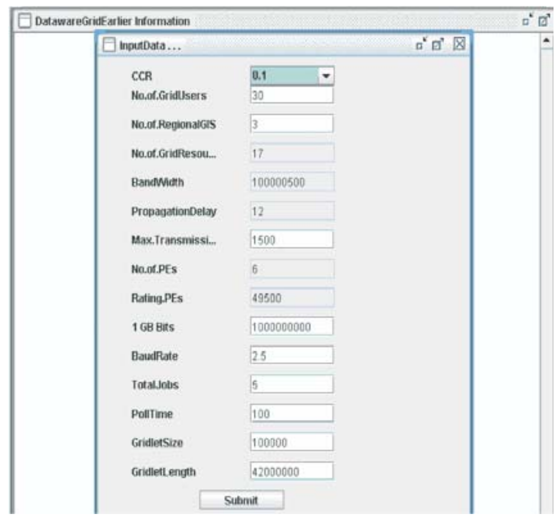


Fig. 2. This picture represents the Normal grid scheduling with the CCR value as 0.1.

The output for the above input is taken from the simulation result as shown in Fig 3

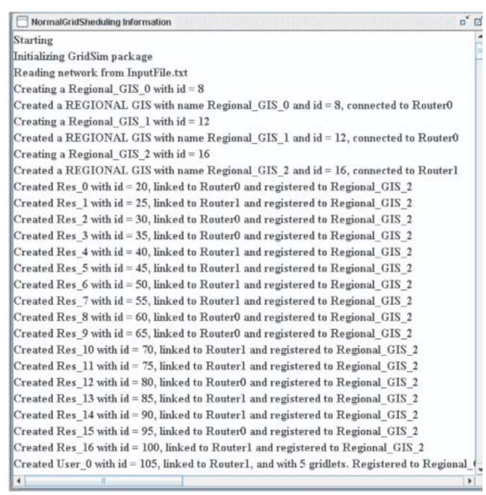


Fig. 3. This Picture gives the simulation result for CCR value 0.1

Like wise for the various CCR values we haven taken the output and comparision is done.

B. Data Scheduling with Prior Error Notification and Rectification

Here we concentrate on how data is transferred with the Prior error detection mechanism and how the errors are rectified. From the simulation result we give the same value for the CCR which indicates communication to computation ratioand the result is taken and shown in Fig 4.

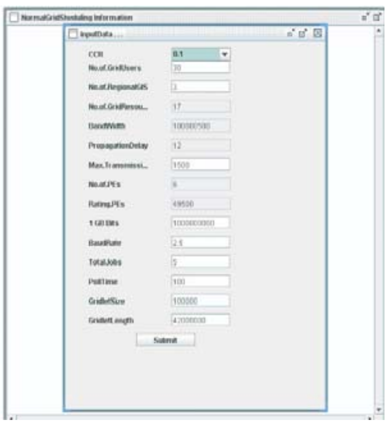
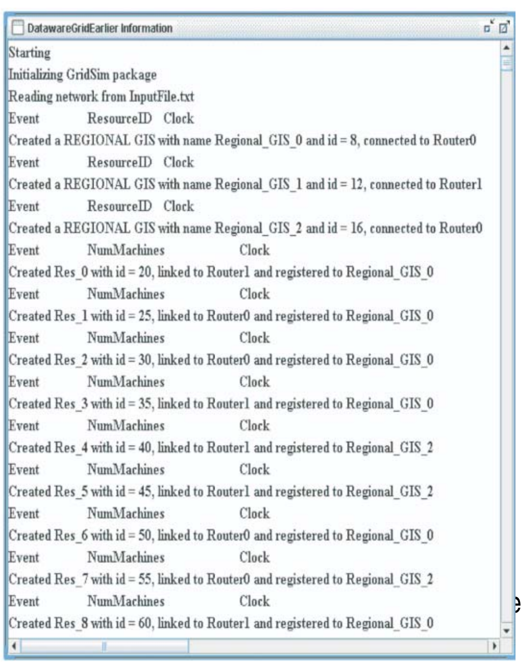


Fig. 4. The picture shows the simulation where an input for CCR is given as 0.1 for the second case.

The input for the CCR is given as 0.1for the data scheduler with prior error notification.The output for the given value of CCR is given in the Fig 5



Different values are given in correspondence to CCR and the resulting values are represented in the Table 1.

Algorithm	X Axis	Y Axis
Data aware grid	0.1	103.99
	0.2	104.99
	0.3	105.98
	0.4	91.99
Data Grid with prior error detection	0.1	90.99
	0.2	95.96
	0.3	96.99
	0.4	91.99

Table 1 The values for x axis and y axis for both data grid and early error rectification.

V. EXPERIMENTS AND RESULT

Based on the simulation results the values were plotted in graph and the result is given in different graphs based on CCR, PPI stands for performance prediction information, and NRR stating normalised resource usage, NM stating normalised makespan.

The first graph shows the execution time in milliseconds for CCR vs Time.The CCR here represents the communication to computation ratio in which both computation time and communication of the node time is together considered for graph and is given in Figure 6.

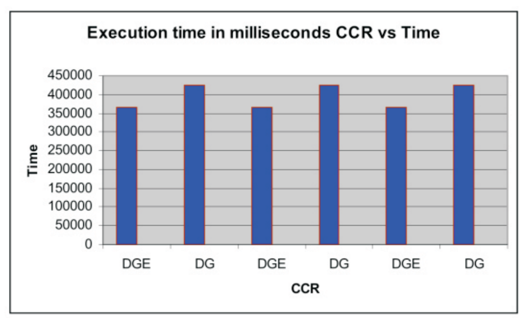


Fig. 6. The graph showing the CCR vs. Time for datagrid (DG) and data grid prior error rectification

Like wise based on the comparison of the performance of both the data grid and the data grid with prior error rectification the execution time of the later is less compared to the former as well the later performs well compared to the data grid .

The normalised value of the resource usage and the performance prediction information is compared in both the cases and is plotted in graph and is given below in figure7 .where the data grid with prior error rectification shows better performance than the data scheduler .the performance of data scheduler is much higher as well the normalised resource usage of the data scheduler also seems to be high in compared the second method

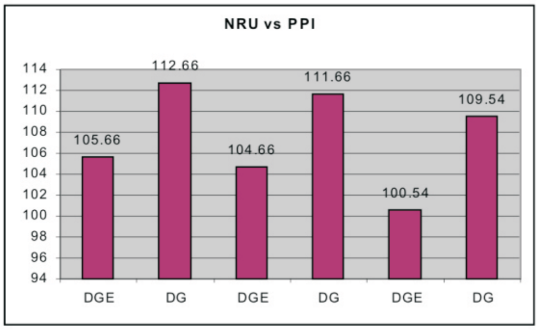


Fig. 7. The Graph showing NRU vs. PPI

VI. CONCLUSION AND FUTURE WORK

In this paper we made a comparative analysis of the data scheduler grid with prior error rectification method and the experiment is implemented in the GridSim and the results are shown in the graph and from the results it is been proved that data scheduled with prior error detection and rectification works more efficiently and fastly than the data grid scheduler .All possible comparisons are made based on CCR, PPI and NRU and based on those only the conclusion is given. The future work will be focused on how the faults are rectified dynamically when the data or the resource is scheduled dynamically to the node in the heterogeneous distributed environment The faults may be like node fault or server problem ,traffic in network and the network itself becoming a problem. All the above problems will be focused and the forth coming work will be enhanced in such manner.

REFERENCES

[1] Radha B. and Sumathy V. "Comparison of ACO and PSO in Grid Job Scheduling" .CIIT International journal of networking and communication Engineering print: ISSN 0974 – 9713 & Online: ISSN 0974 – 9616 DOI: NCE102009003

[2] Kosar T. and Balman M. "A new paradigm: Data-aware scheduling in grid computing. Future Generation Computer Systems", In Press, DOI: 10.1016/j.future. 2008.09.006.

[3] Center for Enabling Distributed Petascale Science. Technical report, A Department of Energy SciDAC Center for Enabling Technology.

[4] Cieslak D., Chawla N., and Thain D. 2008 "Troubleshooting Thousands of Jobs on Production Grids Using Data Mining Techniques". IEEE Grid Computing.

[5] Thain D. and Livny M. 2002 "Error scope on a computational grid: Theory and practice". In Proceedings of the 11th IEEE Symposium on High Performance Distributed Computing (HPDC'02), pages 199–208. IEEE Computer Society.

[6] Kosar T. and Livny M. 2004 "Stork: Making Data Placement a First Class Citizen in the Grid". In International Conference on Distributed Computing Systems.

[7] Mehmet Balman and Kosar T. 2007 "Data scheduling for large scale distributed applications". In the 5th DCEIS, International Conference on Enterprise Information Systems (ICEIS'07).

[8] GridFTP Developer's Guide. http://www.globus.org.

[9] Allcock W., Bresnahan J., Kettimuthu R., and Link M. 2005 "The globus striped gridftp framework and server". In Proceedings of the ACM/IEEE conference on Supercomputing, page 54. IEEE Computer Society.

[10] Antonio Delgado Peris, Jose Hernandez, Eduardo Huedo and Ignacio M Llorente 2010"Data location-aware job scheduling in the grid. Application to the GridWay metascheduler" 17th International Conference on Computing in High Energy and Nuclear Physics (CHEP09) IOP PublishingJournal of Physics: Conference Series 219, 062043 doi:10.1088/1742-6596/219/6/062043

Ms. B. Radha received B.Sc Degree in Chemistry in 2001, MCA Degree in Computer Applications in 2004. She is currently working as a Lecturer in the Department of Computer Applications, Sri Ramakrishna Engineering College, Coimbatore She is currently perusing Ph.D her research interest includes Resource allocation in Grid Computing and Simulation. She has published two technical papers in National conferences and one paper in the international journal. She is a life time member of ISTE.

Dr. V. Sumathy acquired B.E degree in Electronics and Communication Engineering from Government College of Technology (GCT), Coimbatore, in the year 1988. She obtained her Master's degree in Computer Science and Engineering from Government College of Technology, Coimbatore, in the year 2000. She has completed her Ph.D in the area of Ad-Hoc Networks. At present she is an Asst Professor in the Department of ECE, GCT and Coimbatore. She has published nine papers in both National/International Conferences and two papers in journals in the area of Ad-Hoc Networks.