http://dx.doi.org/10.21172/ijiet.113.10

Impact of Cluster Size & Scheduling Policy In Grid Computing

Neelam Bohra[1], Dr. Anil Gupta[2], Shailendra Purohit[3]
[1]*Research Scholar ,Department of Computer Science M.B.M Engineering College, Jodhpur ,Rajasthan*
[2]*Professor and Head, Department of Computer Science and Engineering, MBM Engineering College, Jodhpur*
[3]*Department of Computer Science ACMRC Jodhpur Rajasthan*

Abstract- Grid computing is consider as a evolving technology which exploit geographically separated computing power, .Performance of Grid can be measured in terms of the amount of work they are able to deliver over a period of time In Grid resources are scattered so in order to utilize them completely scheduling plays an important role, the goal of scheduling is to complete the task as early as possible. Among various scheduling techniques, this work include FCFS (First come First serve), EDF (Earliest Deadline First) and Easy Backfilling techniques, these scheduling policies are compared by performance metrics like makespan, tardiness and response time and for better performance the value of these parameters must be minimum In this work we analyses the effect of various cluster (machine) size on above mention scheduling algorithms, both homogeneous and heterogeneous cluster size are consider on the FCFS, EDF and Easy Backfilling scheduling policies and we compare the value of performance metrics in order to find out which algorithm perform well for which cluster size .This work will give an idea to user that which algorithm would be suitable for which cluster size.
Keywords – cluster size, make span, tardiness, response time

I. INTRODUCTION

Grid change the opinion about availability of computing power by merging number of computing device to perform various task In order to find out ideal machine for running jobs submitted by user some kind of scheduling procedure must be used by Grid. In Grid computing scheduling is done at local resource level and application level. In Grid computing various geographical scattered resources are used to solve a problem, Grid need balance use of its resources so that all resources are completely utilizes and waiting time is also minimized. Scheduling plays an important role for efficient use of resources, so resource sharing with appropriate algorithm need for complete task otherwise jobs has to go through for large waiting time .there are so many algorithm already available, in this work we want to analyses the effect of various cluster size on different scheduling algorithm so that overall performance of Grid can be increased. In order to observe the effect of the various cluster (machine) sizes on different scheduling algorithms we have to done experiments in Grid computing. In this work we use Alea2 simulator which is based on GridSim [4]. .

The rest of the paper is organized as follows. Various scheduling phases and types of scheduling in Grid are explained in section II. Experimental results are presented in section III. Concluding remarks are given in section IV.

II. SCHEDULING IN GRID

A. Various Scheduling Phases–

In Grid scheduling concept is used at various layer like O.S, cluster and global system, In Grid system since scheduling is used at various levels, Scheduling can be undergoes in to various stages like to find out the availability of resources, specific requirement of task or job and choosing a scheduling algorithm which decide which type of resources, total number of resources and total duration for which resources granted for completing the end user task can be decided. Figure 3.9 shows the Grid environment which is consist of large virtual organization made up of many Grid site like Gridsite1.... Grid site M, different real organization owned these Grid sites. There is at least one grid level scheduler who is also known as global scheduler or broker whose work is to manage workload but it does not have complete knowledge about local computing resources where application or task has to be processed and most important resources in Grid are independent and they are not completely in control to the Grid scheduler. Generally resources in Grid are independent and local policies of these resources are not disobey by Grid scheduler and each Grid site can have its own Grid scheduler hence there is no central Grid scheduler and more than one VO contain same local resources and each Grid site contain resource scheduler whose work is to schedule and execute the jobs without knowing that they are coming from which Grid application. Resource monitoring infrastructure (RMI) of each Grid site provide necessary information to Grid scheduler which helps in taking better scheduling decision and status of execution of jobs is determined with the help of job monitoring .Figure 3.9 shows the various system component and different scheduling phases in Grid computing ,all these component make possible for Grid computing to utilize the resources properly while consider the factors like their ease of use ,potential, charges and user need etc In Grid there are number of jobs and so many resources are available in such environment the main aim of scheduling in Grid is to find out the best possible way so that these jobs are properly mapped to the available resource[11]

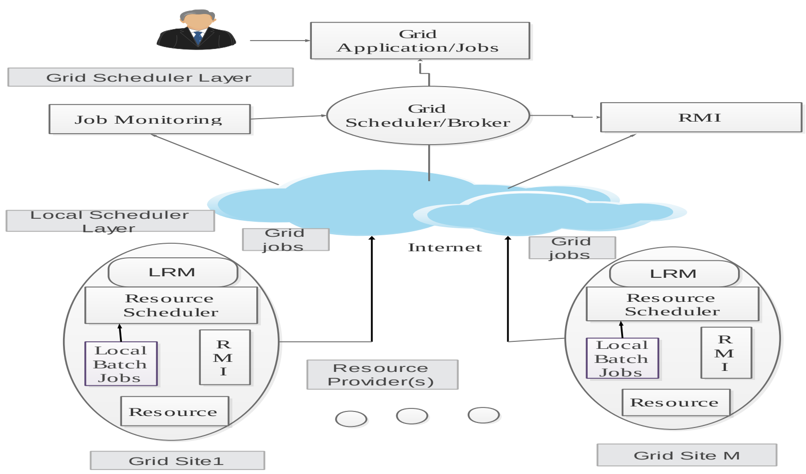

Figure 1 Scheduling Phases in Grid Computing [11}

B. Types of Scheduling in Grid

Scheduling in Grid are concerned at the two level or broadly we can say that there are two types of scheduling exist in Grid computing i.e. one present at the local resource level and another at the application level .

2.1 Resource scheduling

Resource scheduling is done in order to get maximum uses of resources and fair distribution of load on the resource .Resource scheduling further classified as:

1. Time sharing
2. Space sharing

Time sharing scheduling is involved in each machine (computer) of the cluster while LRM uses the space scheduling to schedule the job present in batch queue on free machine available on cluster. There are mainly two type of decision has to be taken in resource scheduling firstly select the jobs from queue and then chose the appropriate node for selected job from existing nodes. These two decision can be taken with the help of various algorithm, figure 3.10 classify the various algorithm which are involved in resource scheduling

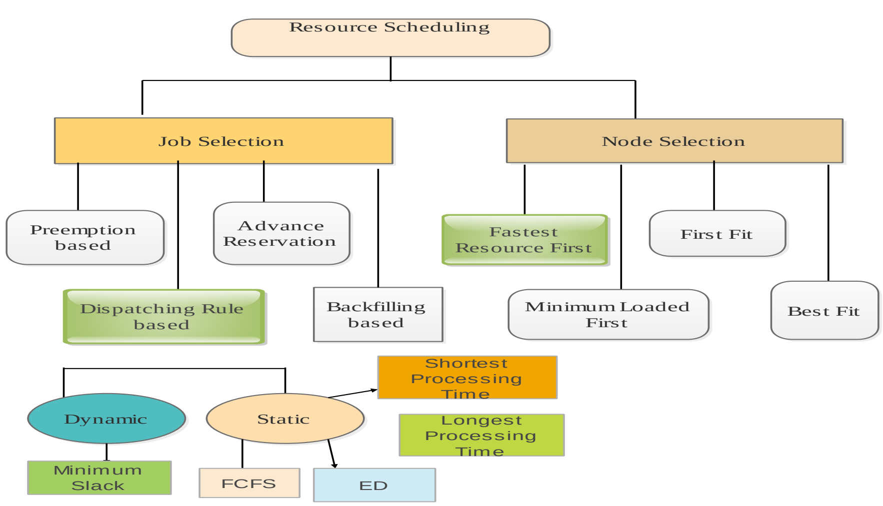

Figure 2.1 Resource Scheduling Algorithm

Following are the job selection algorithms through which decision is taken to pick job from queue of jobs:
a. Preemption based: This resource scheduling algorithm is used to schedule high priority job, when high priority job arrives then resources are taken from low priority job and high priority jobs are start execution
b. Advance Reservation based: This resource scheduling algorithm is beneficial for those users who want to reserve resources for future use, this algorithm assure that jobs will get the required resource when jobs arrives so that jobs need not to stay in queue
c. Backfilling based: This resource scheduling algorithm is extension of FCFS algorithm, if any job in the queue does not get the needed resources then this job has to wait for required resources and it can't execute then job comes after this job start execution without delay
d. Dispatching Rules based: This is most easy resource scheduling algorithm, it can be further classified as 1) Static 2) Dynamic
Following are the Node selection algorithm through which decision is taken to select appropriate node from existing nodes:
In resource scheduling with the help of above algorithm when job is selected then with the various available node algorithm appropriate node is chosen from the available nodes, below some node scheduling algorithms are discussed
a) Best Fit
This algorithm chooses that node which has less number of resources available but it is capable to complete the task.
B)First Fit
This algorithm provide first suitable node from the list of existing node.
c) Fastest Resource First
This algorithm selects that node which contains fastest resources from the available nodes.
d) Min Loaded First
This algorithm choose that node which contain resource having unused CPU power or minimum utilization for completing task

2.2 Application scheduling
Entire application is schedule at Grid scheduler where user submit his task .Grid scheduler is also known as global scheduler, meta scheduler, broker etc Main objective of Grid scheduler is to allocate a Grid site to jobs where resource scheduler assign a machine of cluster to complete the task, for this allocation of Grid sites various application scheduling algorithms are used.

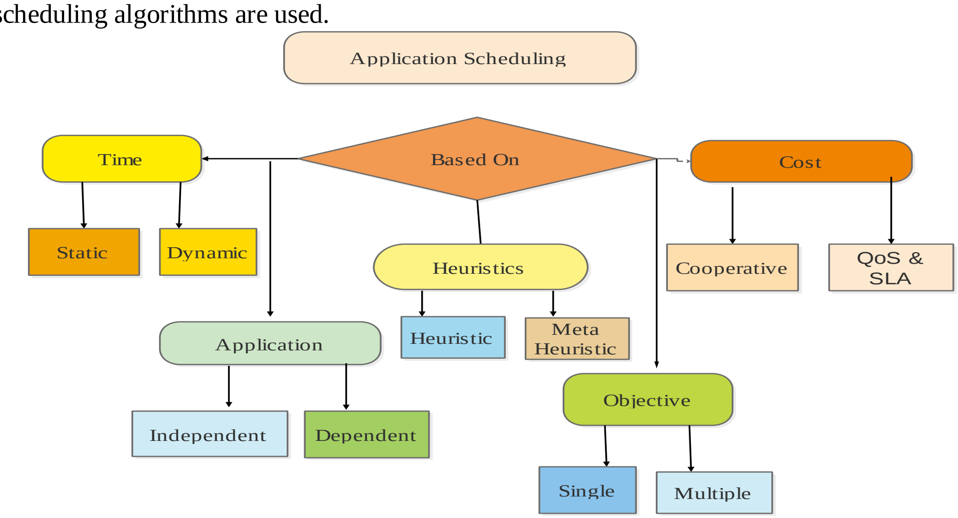

Figure 2.2 Application Scheduling Algorithm

Classification of application scheduling is done on the basis of various factors; some of them are discussed here:
i) Based on Time
In this category two types of scheduling comes Static and Dynamic

a) Static
In this type of scheduling schedule of all the jobs is known before the execution of application is start .It is appropriate for non changeable environment; in such environment it can be easily calculate the execution time of jobs
b) Dynamic
In this there is no prior schedule is known for application, it comes to known only when jobs are ready for execution, This type of scheduling is suitable for dynamic or changing environment ,such kind of scheduling algorithm are used for jobs coming online or real time jobs where execution time of jobs cannot be find out
ii) Based on Application
Application scheduling can be categorized on the basis of application
a) Independent Task
In independent task, all jobs of the application are independent to each other hence they can be schedule in any order
b).Dependent Task
In such kind of application task are dependent on each other and task are ready to execute only when predecessor task finished their execute
iii) Type of Heuristics
It is expected that scheduling algorithm must be feasible and nearly optimal, there are two kinds of Heuristics are present
a) Heuristics
For specific kind of problem such approach is used to design algorithm
b) Meta –heuristics
For specific problem this type of heuristics are used to make a general approach to create algorithm.
iv) Number of Objectives
Scheduling can be classified according to number of objective involve in scheduling decision, according to this there are two types of scheduling
a) Single objective scheduling: In such kind of scheduling, scheduler has to concentrate only on single objective while taking scheduling decision.
b) Multiple objective scheduling: In this kind of scheduling, scheduler has to be consider multiple objective while taking scheduling decision.

v) Type of objective function
On the basis of attribute of objective function scheduling algorithm can further classify into following two categories
a) Minimization based scheduling: If objective are minimum then automatically cost is also minimum, in such kind of scheduling scheduler try to minimize the objective.
b) Constrained based scheduling: In such kind of scheduling, information of the upper bound value objective is provided to scheduler and scheduler has to make schedule algorithm which fulfil this range.
vi) Cost of resource usage
On the basis of whether fee is charged or not for the usage of resource there is following two type of scheduling algorithm present
a) Cooperative/volunteer access based scheduling: In such kind of scheduling there is no charge is taken from resource users, the main goal of scheduler is to reduce the execution time, and number of Grid projects uses such kind of scheduling.
b) QoS and SLA based scheduling: This type of scheduling is also known as utility based
Grid computing in this type of scheduling user has to pay to the owner of resources for the usage of resources, hence in this monetary cost is deciding factor for scheduling

III. EXPERIMENT AND RESULT

This work uses Alea2 the extended version of GridSim to simulate the scheduling process in a grid computing environment. By these experiments we are comparing three different scheduling policies FCFS, EDF and Easy Backfilling. All experiment are perform on real dataset HPC2N Seth log by keep the number of jobs fixed i.e. 3000, we vary the size of cluster but total no of processor remain same these experiments are perform at the constant CPU speed and RAM per node is also having fixed value i.e. 1048576KB. We make a change in only cluster id, cluster name, number of nodes and CPU per node in machine files which is created manually by us

3.1 Impact of various Homogeneous clusters on scheduling policies

In this experiment we use homogeneous cluster of various size

a. 240 x 1, b.60 x 3, 60 x 1, c.30 x 4, 30 x 3, 30 x 1

In this experiment there are three scenarios, in each scenario we apply FCFS, EDF and Easy Backfilling on these above mention cluster size and obtain the value of makespan, tardiness and response time and finally we compare these entire three algorithms together for these performance metrics

SCENARIO 1 For FCFS Scheduling Policy

Table3.1 Comparison of performance metrics value for homogeneous clusters on FCFS

FCFS	240X1	60X3,60X1	30X4,30X3,30X1
Makespan	5121737	5127972	5392225
Tardiness	1584.37	2557.07	20839.97
Response Time	13747.68	15527.77	37040.11

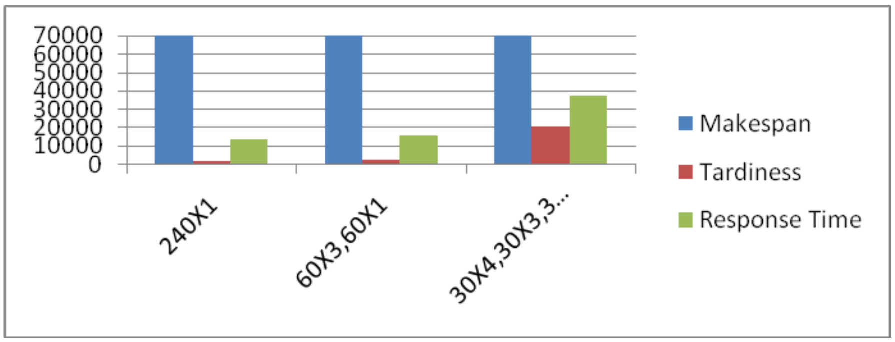

Figure 3.1 Comparison of evaluation criteria for various homogeneous cluster sizes in FCFS

SCENARIO 2 For EDF Scheduling Policy

Table3.2 Comparison of performance metrics value for homogeneous clusters on EDF

EDF	240X1	60X3,60X1	30X4,30X3,30X1
Makespan	5126163	5139170	5305778
Tardiness	586.25	1143.26	11192.79
Response Time	12905.64	13961.31	26966.72

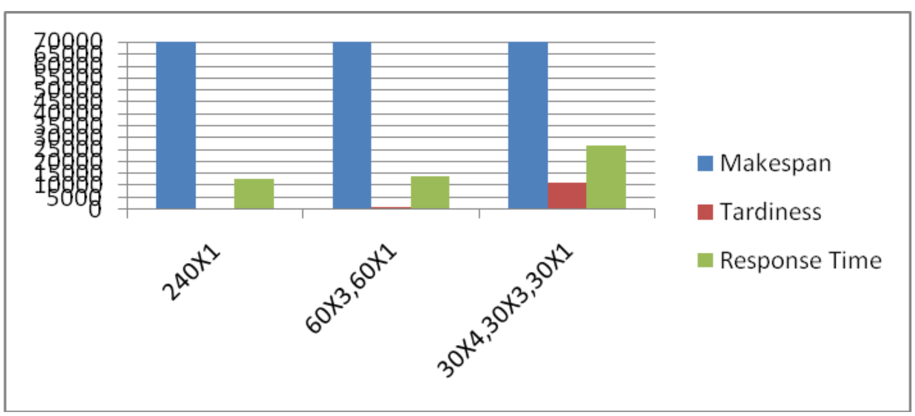

Figure 3.2 Comparison of evaluation criteria for various homogeneous cluster sizes in EDF

SCENARIO 3 for Easy Backfilling scheduling Policy

Table 3.3 Comparison of performance metrics value for homogeneous clusters on Easy Backfilling

Easy Backfilling	240X1	60X3,60X1	30X4,30X3,30X1
Makespan	5120855	5120855	5120855
Tardiness	1082.94	1233.48	1878.44
ResponseTime	12566.28	12736.87	13670.09

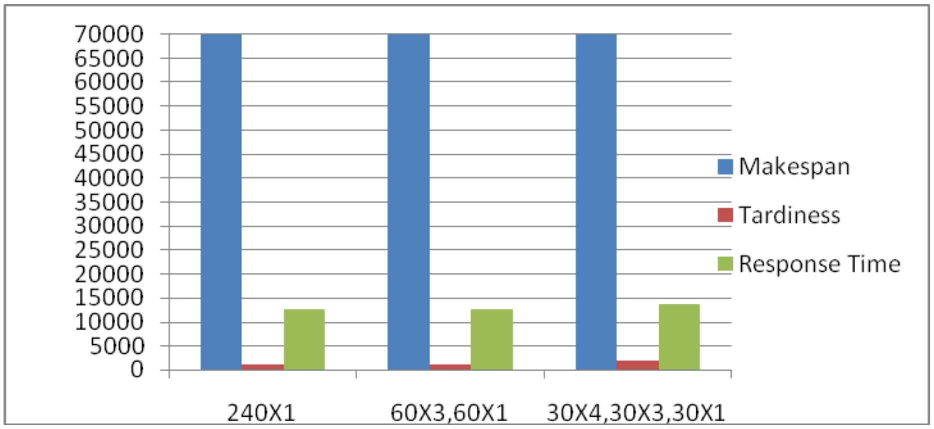

Figure 3.3 Comparison of evaluation criteria for various homogeneous cluster sizes in Easy Backfilling

3.2 Impact of various Heterogeneous clusters on scheduling policies

In this experiment we use following cluster of various size

a) 240 x 1 , b) 120 x 1 , 60 x1 , 30 x 2 , c) 60 x 1 , 40 x 3 , 30 x2

In this experiment there are three scenarios, in each scenario we apply FCFS, EDF and Easy Backfilling on heterogeneous clusters and obtain the value of makespan, tardiness and response time and finally we compare these entire on three algorithms together .

SCENARIO 2(a) For FCFS Scheduling Policy

Table 3.4 Comparison of performance metrics value for Heterogeneous clusters on FCFS

FCFS	240x1	120x1,60x1,30x2	60x1,40x3,30x2
Makespan	5121737	5318933	5318933
Tardiness	1584.37	18371.66	18371.66
Response Time	13747.68	34472.64	34472.64

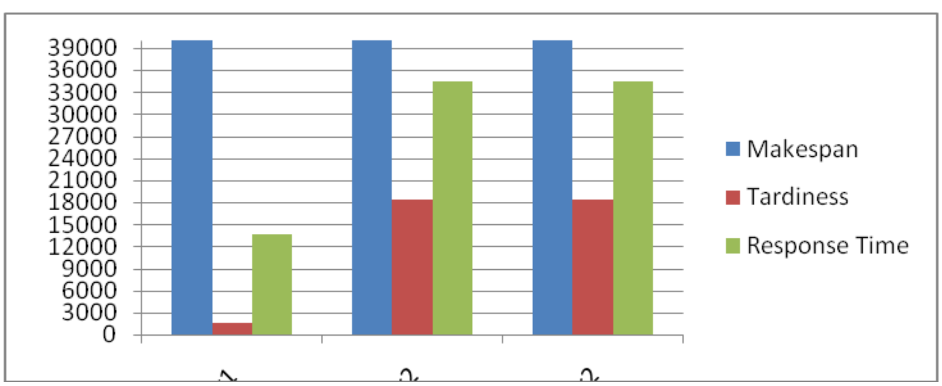

Figure 3.4 Comparison of evaluation criteria for various heterogeneous cluster sizes in FCFS

SCENARIO 2(b) For EDF Scheduling Policy

Table 3.5 Comparison of performance metrics value for Heterogeneous clusters on EDF

EDF	240X1	120x1,60x1,30x2	60x1,40x3,30x2
Makespan	5126163	5307726	5493524
Tardiness	586.25	12353.59	15015.09
Response Time	12905.64	28632.6	32107.98

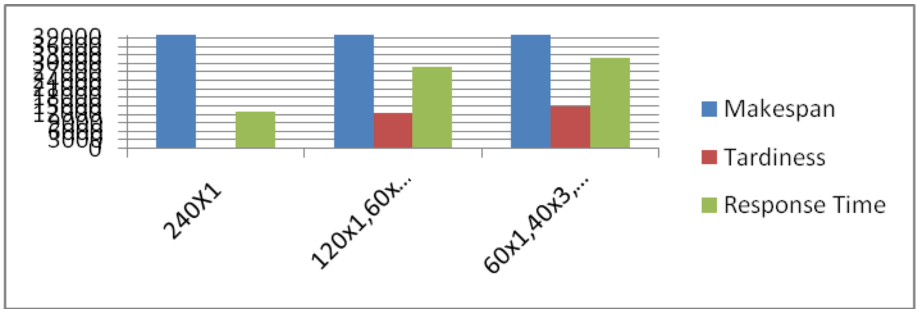

Figure 3.5 Comparison of evaluation criteria for various heterogeneous cluster sizes in EDF

SCENARIO 5(c) For Easy Backfilling scheduling Policy

Table 3.6 Comparison of performance metrics value for Heterogeneous clusters on Easy Backfilling

EDF	240X1	120x1,60x1,30x2	60x1,40x3,30x2
Makespan	5120855	5120855	5282046
Tardiness	1082.94	1857.79	2203.51
Response Time	12566.28	13537.71	14175.41

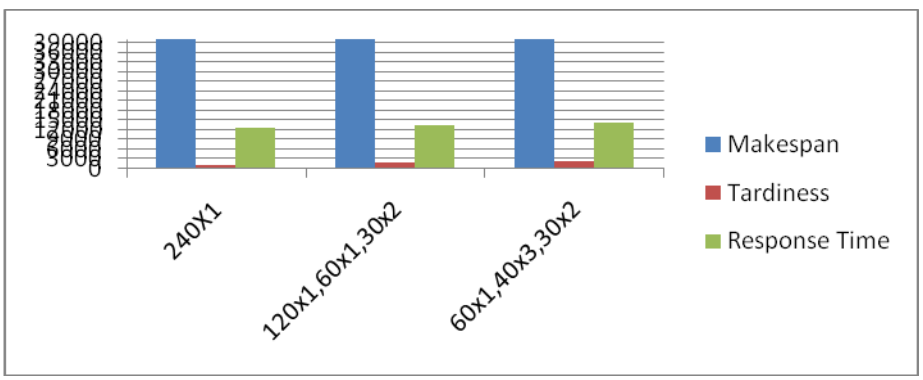

Figure 3.6 Comparison of evaluation criteria for various heterogeneous cluster sizes in Easy Backfilling

IV. CONCLUSION

In these experiment there are three scenarios in each scenario we apply FCFS, EDF and Easy Backfilling algorithm on homogeneous and heterogeneous clusters and obtain the value of makespan, tardiness and response time, for better performance it is expected that value of all these three parameter must be minimum, from the obtained values we conclude that

4.1. For FCFS algorithm

The impact of homogeneous clusters on FCFS is that as the size of node increases the value of all the three parameter is also increases and for heterogeneous cluster again value of all the three parameter is increases with increase in node size,it is also observe that value of all the parameter is higher in homogeneous clusters compare to heterogeneous clusters .

4.2 .For EDF Scheduling Policy

The impact of homogeneous clusters on EDF is that as the size of node increases the value of all the three parameters is also increases and in case of heterogeneous clusters increase in node results increase of all the three parameter but compare to heterogeneous clusters value of all the three parameter in homogeneous clusters are minimum.

4.3. For Easy Backfilling Scheduling Policy

In this scheduling policy in case of homogeneous cluster the value of all the three parameter increases slightly as the number of node increases and same hor the heterogeneous clusters value of all the three parameter increases with increase in size of node but compare to heterogeneous clusters the value of these parameter in homogeneous clusters is less

So it is concluded that as the number of nodes increases value of performance criteria is also increases which is not desirable and in case of EDF and Easy Backfilling Scheduling Policy the homogeneous clusters perform better compare to heterogeneous and while in case of FCFS value of these performance metrics is lower in heterogeneous cluster than homogeneous clusters.

V. REFERENCE

[1] Ian Foster Carl Kesselman § Steven Tuecke" The Anatomy of the Grid Enabling Scalable Virtual Organizations" in the International Journal of High Performance Computing Applications, 2001

[2] Volker Hamscher1, Uwe Schwiegelshohn1, Achim Streit2, and Ramin Yahyapour"Evaluation of Job-Scheduling Strategies for Grid Computing" in Springer-Verlag Berlin Heidelberg 2000

[3] Bart JacobLuisFerreira, NorbertBieberstein, CandiceGilzean, JeanYvesGirard, RomanStrachowskiSeong "Enabling Applications for Grid Computing with Globus" (Steve) Yu (June 2003)

[4] Dalibor Klusáˇcek and Hana Rudov"Alea 2 – Job Scheduling Simulator" Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques in 2010)

[5] Indu Gandotra, Pawanesh Abrol, Pooja Gupta, Rohit Uppal and Sandeep Sing h"Cloud Computing Over Cluster, Grid Computing: a Comparative Analysis" in Journal of Grid and Distributed Computing, 2011

[6] Priti Lohani 159.735 Final ReportCluster Scheduling Submitted

[7] Zafril Rizal M Azmi1, Kamalrulnizam Abu Bakar2, Mohd Shahir Shamsir3, Wan Nurulsafawati Wan Manan4, Abdul Hanan Abdullah"Scheduling Grid Jobs Using Priority Rule Algorithms and Gap Filling Techniques" in International Journal of Advanced Science and Technology December, 2011

[8] Reza Fotohi and Mehdi Effatparvar " A Cluster Based Job Scheduling Algorithm for Grid Computing "I.J. Information Technology and Computer Science, 2013

[9] Krzysztof Kurowski ,Ariel Oleksiak,Wojciech Pia˛tek , Jan W˛eglarz"Hierarchical scheduling strategies for parallel tasks and advance reservations in grids "Journal of Scheduling4 2013

[10] R. Joshua Samuel Raj Kalasalingam University "Intelligence in Scheduling for Grid Computing" in Kalasalingam Academy of Research and Education in January 2014

[11] Harshadkumar B. Prajapati and Vipul A. Shah"Scheduling in Grid Computing Environment" (IEEE- 2014)

[12] Sandip Fakira Lokhande1, Sachin D. Chavhan2, Prof. S. R. Jadhao"Grid Computing Scheduling Jobs Based on Priority Using Backfilling" International Journal of Electrical Electronics & Computer Science Engineering ,2015

[13] Bimal VO and M. Anand Kumar" Performance Analysis of TimeLine Algorithm against CONS, PBS_PRO and BestGap in Grid Environment using Alea " International Journal of Applied Engineering Research,2016

[14] Bimal VO and M. Anand Kumar"Cluster usage analysis of timeline based grid scheduling algorithm" International Journal of Computer Trends and Technology (IJCTT) ,2017

[15] Carsten Ernemann, Volker Hamscher, Uwe Schwiegelshohn, Rami Yahyapour"On Advantages of Grid Computing for Parallel Job Scheduling" in Cluster Computing and the Grid, 2002. 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID.02)

[16] Carsten Ernemann , Volker Hamscher, Ramin Yahyapour"Benefits of Global Grid Computing for JobScheduling" in Proceedings of the 5nd IEEE/ACM International Symposium on Cluster Computing and the Grid, 2004

[17] Book written by Frederic Magoules Jie Pan Kiat" Introductiobbn to grid computing" International Standard Book Number 13:978-1-4200-7407-9

[18] Berman, Geoffrey Fox, and Tony Hey "The Grid: past, present, future" Grid Computing – Making 2002 ISBN: 0-470-85319-0

[19] Book written by Ian Foster, Carl Kesselman "The Grid 2: Blueprint for a New Computing Infrastructure"

[20] Mark Baker ,Amy Apon Rajkumar Buyya , Hai Jin" Cluster Computing and Applications"in 18th September 2000

[21] Dalibor Klusáček and Hana Rudová. Alea 2 - Job Scheduling Simulator. In proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques (SIMUTools 2010), ICST, 2010.

[22] Workload traces uses in this work are available either at my page or at Parallel Workloads Archive.