Handling Complex Use Scenarios for Selection of Best Resources in Grid Environment

[1]Lata Gadhavi , [2]Madhuri Bhavsar
Institute of Diploma studies
Nirma University
Ahmedabad, Gujarat, India
[1]lata.gadhavi@nirmauni.ac.in,[2]madhuri.bhavsar@nirmauni.ac.in

Abstract

A Grid computing system is comprised of large sets of heterogeneous and geographically distributed resources that are aggregated for executing applications.As the number of resources in Grids increases rapidly, selecting appropriate resources for jobs has become a crucial issue.To avoid single point of failure and server overload problems, use scenario provides an alternative means of resource selection in distributed systems.Here decentralized broker selects computational resources based on actual job requirements, job characteristics, and information provided by the resources through webmds. Based on the previous history of the resources and performance, we can specify that which resource is best from all the available resources then run the job on that resource which fulfills the requirements of the job.

Index Terms:-*Performance-based Selection of Best Resources,Grid Computing.*

1 Introduction

Grid Computing is an active research area which promises to provide a flexible infrastructure for complex, dynamic and distributed resource sharing system.Globus a middleware which is a de facto standard for grid computing.This scenario contains setup for a grid environment, in which 50s or 100s of execution nodes out of which one has been choosen as a submission node, and the other two are container nodes. These container nodes are actually having the schedulers of OpenPBS, SGE and also condor which creates heterogeneous environment for clusters. When the submission node submits jobs to the containers, then these jobs are scheduled to run on other nodes attached to the OpenPBS and SGE clusters in parallel.Scenarios are ordered by increasing complexity,where complexity is loosely defined as the number of decisions that must be made by the GRID S/W. Handling complex use scenario contains the Immediate job execution[1],which selects the best resource from the all available resources.

2 Related Work

Ivan Roderoa,Francesc Guimb,Julita Corbalan [2] has defined The increasing demand for high performance computing resources has led to new forms of collaboration of distributed systems, such as grid computing systems. Moreover, the need for inter operability among different grid systems through the use of common protocols and standards has become increased in the last few years.In particular, they consist of the proposed "best Broker Rank" broker selection policy and two different variants. The first one uses the resource information in aggregated forms as input, and the second one also uses the broker average bounded slowdown as a dynamic performance metric.
Oliveira & Lechuga,[3] has defined the pricing of

resources on computational grids where In order to determine a price for the resource's usage,first it is necessary to identify the target of consumption, i.e.,what is being ordered. Following the logic that an outstanding characteristic of the grid is heterogeneity ,it is known that a certain job may have different execution time depending on the peer to which it is submitted. However,the size of the job remains the same.The proposed model of pricing establishes a direct relationship between a unit of the processing metric(PMU), which is the smallest part to which the tariff is applied, and the quantity of instructions executed by the job in the resource,i.e,its size.
Erik Elmroth & Johan Tordsson[4] has defined The task of a Grid resource broker and scheduler is to dynamically identify and characterize the available resources, and to select and allocate the most appropriate resources for a given job. The resources are typically heterogeneous, locally administered, and accessible under different local policies. A decentralized broker, as the one considered here, operates without global control, and its decisions are entirely based on the information made available by individual resources and index servers providing lists of available resources and aggregated resource information.

3 Proposed Algorithm for Best Resource Selection

3.1 Introduction of the algorithm

Proposed Resource selection algorithm will perform a series of tasks, e.g., it processes the specifications in the job requests, discovers and characterizes the resources available, and performs the actual job submission. This algorithm presents a high-level outline of the tasks which are performed in grid environment. The input task specification(s) contains one or more job requests including information about the application to run (e.g., executable, arguments, input/output files), actual job requirements (e.g., amount of memory needed, architecture requirements, execution time required,million instructions per second(MIPS),Gflops), and optionally, job characteristics that can be used to improve the resource selection. The user can include a request for advance reservations.Steps of algorithm are shown in below[2]

- Step 1 , the user's request is processed and split into individual job requests.
- Step 2, the broker discovers what resources are available by contacting one or more index servers & by the Webmds.
- Step 3,The specific characteristics of the resources which are found in the grid environment,are identified through configuration & configuration of the resource list, by querying each individual resource. Each resource may provide static information about architecture type, memory configuration,CPU clock frequency, operating system,clock rate,etc., and dynamic information about current load, batch queue status and various usage policies. Steps 2 and 3 are both performed by LDAP queries sent from the broker to the index servers and the resources, respectively.
- step 4, The actual brokering process is mainly performed here, which is repeated for each job request.
- Step 5,The loop is repeated until either the job is successfully submitted or all submission attempts fail, the latter causing the job to fail.After the configuration of the all available resources then it shows the information about the resources which are connected in the Grid. First of all the demand of resource will get by using some configuration & by using the gprof functionality . It will show the MIPS of the task & execution time required for the job & kflops of the job & resource. To comparing job & node parameters,it will select best resource from all the available resources which is shown in the comparison table.

- Step 6,first job will submit on main node. When the requirements of the job will match with available resources then it will select that resource and run job on that resource which is the Best resource,otherwise job will run on all ordinary resources which are available.
  Based on the previous history of the resources and based on performance measured by the configuration user can specify that which resource is best from all the resources then run the job on that resource which fulfills the requirements of the job. By comparing the job & all available resource's kflops,here got that resource which resources kflops is best suitable with the demanding kflops so by using the programming on the grid it will select the Best resource from the all available resources.

3.2 Pseudocode of the algorithm

- Algorithm- Job submission and selecting resource.
  Input:-Job Description,Resource Information
  Action: Job identifier(s) for the submitted job(s) & searching the best resources from the available resources.
  Output: Run job on the best resource from all available resources based on the performance.
  Job identifier(s) for the submitted job(s). Create a list of all individual job requests.
  Contact one or more index servers to obtain a list of available clusters.Query each resource for static and dynamic resource information(hardware and software characteristics, current status and load, etc).

```
for each job do
       Filter out resources that do not
       fulfill the job requirements.
       Use best fit algorithm for
       matching criteria.
   If get the best match as per requirement.
      Submit the job to the selected resource.
   else
      Submit job on randomly selected
      resource.
   end if
end for
```

Process of best resource selection through algorithm is shown in the figure.

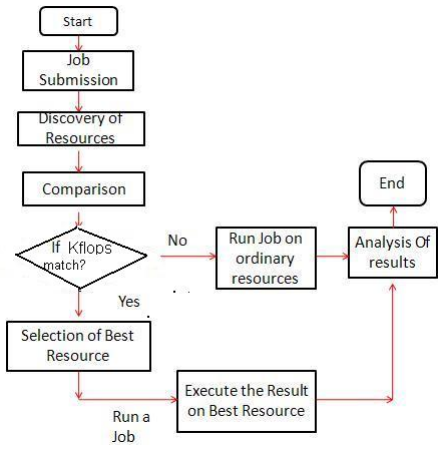

Figure 1: Flow Graph for Best Resource Selection

4 Handling Complex Use Scenario

4.1 Immediate job execution

- Run a job on a specified grid computer:- In a basic approach to remote job submission,the user specifies the execution host to be used and submits a job for which either the code already exists on the target machine or else is uploaded as part of the request.
- Run a job on Best grid computer:- Instead of mandating the specific grid computer in the request ,a user may run a job on the "best" computer defined to the host that is based on the comparison of available and required resources. To select the Best Resource from all available resources here some implementation steps are done[5].
- Criteria for the Best Resource is applied that is Quickest & Cheapest resource,based on

the previous history of the resource's performance,then with the comparison of job's parameter requirement for resource and available resources'parameter. Here Comparison is held & got the best resource from all available resources.Implementation for getting available resources is done using the Java Programming. First here described the algorithm for selection of best resource,then process of job execution & implementation steps of algorithm.

5 Implementation

5.1 Globus Configuration

- Globus is a grid middleware, considered to be the de facto standard for grid computing. So, for practical experience of the grid computing it is best to start with the installation of Globus.
- Operating System : Fedora 8.0
- Middleware : Globus 4.2.1
- Nodes for grid environment-

  guser01-nodeB.grid.nirma.com(Ip-10.1..3.13)(server node)
  guser02-nodeA.grid.nirma.com(Ip-10.1..3.14)(client node)
  guser03-nodeC.grid.nirma.com(Ip-10.1..3.19)(PBS node)

5.2 Process of Job Execution

- Computational grid which explores idle processing power is formed in the lab.This heterogeneous grid environment is configured by the interconnecting 25 nodes with GT 4.2.1 & condor 7.2.1.
- Head node is loaded with globus & condor,where jobs are submitted by the user.Other nodes are configured with PBS & SGE.

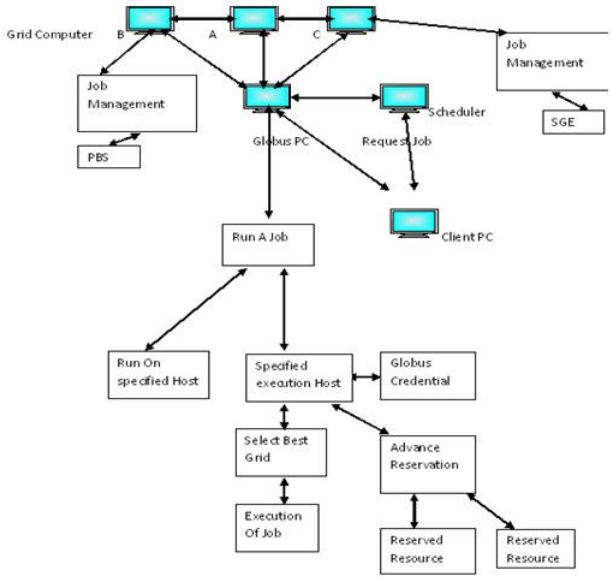

Figure 2: Process Of Job Execution

- Job will submitted on head node and then it will run on the grid and it will show the result of the job execution.Flow of an algorithm as shown in fig.1 & fig.2 ,which elaborates the process of job execution.
- **Submission of Job:-** Here new user will submit job on the main node & when that job will submitted successfully then that job will compare the parameters of jobs with available resources.Here all available resource's list is shown in the table.In the table it shows the all available resource's parameters.

5.3 Requirement of resources for Job

- After the submission of job here criteria of job will compared with resources's parameters. When the requirements of the job will match with available resources then it will select that resource and run on that resource which is

Table 1: Available Resources

Machine Name	Load average	OS	Kflops	MIPS
pc-1	1.000000	WINNT51	1307936	4421
pc-19	0.000000	WINNT51	1349050	5222
Grid2	0.000000	WINNT51	586644	1655
mtech-1	0.120000	WINNT51	1386172	2655
	0.13		3452662	13953

matched with required criteria then Best resource otherwise job will run on all ordinary resources. Based on the historical performance of resources and based on configuration detail identified by webmds,user can specify that which resource is best from all the resources and run the job on that resource which fulfills the requirements of the job.First i obtained the existing job requirements for running on the resource. It shows the parameters related job which covers job location,Job type,MIPS,kflops(floating point instructions per second),time taken for execution ,etc. It is shown in the tableI. Steps for job requirements are,
(1) Availability of resources is checked against the required resources.
(2) Then algorithm selects the best resource to satisfy the need of job.
Requirements of Job properties is shown in the table.

Table 2: Requirement of resources for Job

No	Job Location	Job Type	Kflops	Time(sec)
1	D:/bc/bin	tsp.c	140	0.01
2	D:/bc /bin	mm1.c	60	0.03
3	D:/bc/bin	Performance.c	100	0.02

5.4 Comparison between job & available resources's parameters

- Here comparison is held based on the kflops criteria.Programs kflops & Available resources kflops are compared.By comparing the kflops got the results for Best resource ,which resources kflops is best suitable with the demanding kflops so by using the programming on the grid it will select the Best resource from the all available resources.
As shown in the figure Best resource is selected amongst all the resources.After the selection of the Best resource job will only run on that resource,job will execute on that resource & it will give better result than the ordinary resources. If job will run on Best resource then there are execution time to execute job will be taken less comparatively other resources because Best Resource is Quickest & faster than other resources in grid environment.

Machine Name	Load average	OS	Kflops	MIPS	Total Load Avg
Slot1@A208cc-pc-1	1.00000	WINNT51	1307936	4421	1.100000
Slot1@A208cc-pc-19	0.000000	WINNT51	1349050	5222	0.000000
Grid2.nitdomain	0.000000	WINNT51	586644	1655	0.000000
Slot1@mtech-1	0.120000	WINNT51	1386172	2655	0.120000
	1.12		4588688	13953	1.22

Figure 3: Comparison between demand of job's Kflops & Available Resource's Kflops

5.5 Comparison Results

In this table it shows the all resources which are available and kflops of the resources & selected resource for job submission. It will find the which resource's kflops match with the job's requirements, which calls kflops of job and other functionality parameters related with the job,which is submitted on the main node.

6 Analysis of Results

After comparison for the job requirements and available resources we can analyze that amongst the all resources which resource is best resource.Then job will submit on that resource and execution of job will held on that resource & it gives the results for running job on selected resource.

7 Conclusion

It is envisaged that the grid infrastructure will be a large-scale distributed system that will provide high-end computational and storage capabilities to differentiated users.Complex scenario appeared while job submission & execution are identified.Selection of best resource is identified in the implementation phase. Resource brokering algorithm finds a suitable & best match in the existing grid environment.It also obtains the demanding resources & compares with available resources. In the conclusion,it is observed that best grid nodes improves the performances of the job.

References

[1] Daniel. Minoli, A networking approach to grid computing, Wiley Publications, 2006

[2] Ivan Roderoa,Francesc Guimb,Julita Corbalan,Grid broker selection strategies using aggregated resource information,Computer Architecture Department, Technical University of Catalonia (UPC),Spain,2009.

[3] Oliveira, J. F. A. & Lechuga, T.A,Pricing of resources on computational grids using Globus ,Faculdade Ruy Barbosa, Salvador, Bahia, Brazil. February,2007.

[4] Erik Elmroth,Johan Tordsson,Department of Computing Science,Resource selection based on performance predictions, Umea University,Sweden ,June 2007.

[5] Wu-Chun Chung, Ruay-Shiung Chang,Department of Computer Science, "A new mechanism for resource monitoring in Grid computing",National Dong Hwa University, Hualien, Taiwan,June 2008