A SYSTEM AND METHOD FOR MONITORING AND MANAGING RESOURCES IN A GRID COMPUTING ENVIRONMENT

Field of the Invention

[0001] The present invention generally relates to job scheduling for a grid computing system. In particular, the invention relates to a method and system for monitoring and managing resources in a grid computing environment.

Background of the Invention

[0002] Grid computing systems comprise of several resource nodes in a networked system. Each resource node may further comprise of a single machine or several machines forming another network. By utilizing the computing power of all these machines in this grid computing system, calculations that usually require super computers may be realized by the combined computing capability of all the machines in the system.

[0003] Presently, grid computing has become one of the leading ways for the sharing of resources and computing power over the internet. Several resource nodes of a grid computing system may be able to utilize each other's spare or idle computing power. An example of such sharing is when a resource node is located in a differing time zone, say at night where the computing power of the machines in that node are idle. Another resource node currently requiring extra computing power may be able to send jobs to this idle node to speed up the calculations or even to free up computing power in its own node for more crucial applications.

[0004] With present grids spanning global proportions, the monitoring, scheduling and assigning of jobs to nodes have become a real concern. To efficiently tap into idle resources of these resource nodes, systems and methods had to be developed to

keep track of progress of the member nodes executing the jobs as well as available nodes for assigning jobs to. The tracking and monitoring becomes more and more difficult as the nodes in the grid increase and span such global proportions. Present tracking and monitoring methods also disadvantageously taps into the computing power that may well be used in the execution of other jobs.

[0005] An example of which is disclosed by Bantz in US Patent Application Publication No. 2004/0003077A1. Bantz discloses a method for allocation of grid computing to network stations by first characterizing the availability of the workstation resources. The characterization is performed by real time monitoring and analyzing of the CPU utilization of the workstation. Grid workloads or jobs are then scheduled on workstations characterized as having low CPU utilization.

[0006] Furthermore, present methods for assigning jobs to nodes are based on a round robin queuing system that is disadvantageously arbitrary and takes no consideration of the characteristics of the idle or available resource nodes.

[0007] It can thus be seen that there exists a need for a simple and robust method and system for monitoring and managing grid resources in a grid computing environment.

Summary of the Invention

[0008] The present invention seeks a method and system for monitoring and managing grid resources in a grid computing environment.

[0009] Accordingly, in one aspect, the present invention provides, a method for monitoring and managing grid resources comprising of a plurality of resource nodes

and at least one supervisor node in a grid computing environment, the method comprising:

a) receiving a job;

b) performing job brokering to match resource node to job by a job broker of the supervisor node;

c) submitting job to recommended resource node by a job submitter of the supervisor node; and

d) checking status of and managing of all current jobs by a job monitor of the supervisor node.

[0010] In another aspect, the present invention provides, a system for monitoring and managing grid resources comprising of a plurality of resource nodes in a grid computing environment, the system comprising:

an attribute repository for storing attributes and status of the plurality of resource nodes; and

at least one supervisor node having:

a resource broker for retrieving attributes and status from the attribute repository, receiving at least a first and a second priority attribute, and matching the first and second priority attributes to the plurality of resource nodes and recommending at least one matching resource node in respective priority to a job submitter;

the job submitter for submitting a job to one of the plurality of resource nodes in accordance with recommendation by the resource broker.

Brief Description of the Drawings

[0011] A preferred embodiment of the present invention will now be more fully described, with reference to the drawings of which:

[0012] FIG.1 illustrates a Grid computing environment in accordance with the present invention;

[0013] FIG.2 illustrates a flowchart for a method for monitoring and managing grid resources in a grid computing environment in accordance with the present invention;

[0014] FIG.3 illustrates a flowchart for the performing job brokering step of FIG. 2;

[0015] FIG.4 illustrates a flowchart for the checking and managing step of FIG.2; and

[0016] FIG.5 illustrates a flowchart for the performing job brokering excluding failed resource nodes step of FIG.4.

Detailed description of the Drawings

[0017] A method and system for monitoring and managing grid resources in a grid computing environment with a preferred embodiment is described. In the following description, details are provided to describe the preferred and the alternate embodiment. It shall be apparent to one skilled in the art, however that the invention may be practiced without such details. Some of the details may not be described at length so as not to obscure the invention.

[0018] Referring to FIG.1, the present invention makes use of a supervisor node 20 among several resource nodes 30 in a grid computing environment 10. The supervisor node 20 provides services for the monitoring and managing of the

resource nodes 30. The supervisor node 20 may also be considered to be one of the resource nodes 30. In addition to the infra-structural modules that run the supervisor node 20 either as a network of machines or an individual machine, the supervisor node 20 further comprises a resource broker 23, a job submitter 25 and a job monitor 27.

[0019] The resource broker 23, job submitter 25 and job monitor 27 may comprise of software modules running either on the network of the supervisor node 20 or the individual machine that makes up the supervisor node 20.

[0020] The resource broker 23 serves to retrieve information stored in a attribute repository 29 about the resource nodes 30. All the resource nodes 30 would send and update these information on the attributes and status of the resource nodes 30 to the attribute repository 29 periodically. The attribute repository 29 may reside in any of the resource nodes 30 or may be a separate node in the grid computing environment 10.

[0021] The information retrieved about the resource nodes 30 may be predetermined and may comprise of some attributes of the resource node 30 as well as the status and condition of the resource node 30. Some of the attributes that may be available are CPU utilization, Random Access memory (RAM) capacity or Virtual Memory (VM) Capacity. These information are retrieved and consolidated by the resource broker 23 and used to recommend at least one resource node 30 to particular submitted jobs having priority attributes that are matching.

[0022] The job submitter 25 serves to push or send jobs to the resource node 30 as recommended by the resource broker 23. The job submitter 25 also works together with the resource broker 23 to ascertain whether a recommended resource node 30 is available and if not, which other resource node 30 to push the job to.

[0023] The job monitor 27 serves to monitor the status of the jobs being performed at the various resource nodes 30. The job monitor 27 works together with the job submitter 25 to manage the jobs being performed. If a job fails at a particular resource node 30, the job may be resubmitted to another resource node 30. The job monitor 27 will then inform the submitter of the job if the job is completed or if a failure has resulted after a predetermined number of tries.

[0024] The method 200 for monitoring and managing resources in a grid computing environment starts with the step where new jobs may be submitted 210 to the grid computing environment 10. Where there are new jobs submitted 210, the method 200 continues to the step of performing job brokering 215 where at least one resource node 30 may be recommended by the resource broker 23 for running the submitted job.

[0025] The submitted job is then submitted 220 to the recommended resource node 30 for execution by the job submitter 25. The job submitter 25 pushes jobs to the resource node 30 as recommended by the resource broker 23.

[0026] Thereafter, the method 200 continues to the step of checking 225 the status and managing of all current jobs in the grid computing environment 10.

[0027] Referring to FIG.3, the step of job brokering further starts with the step of checking the attribute repository 29 for attributes and status for all resource nodes 30 in the grid computing environment 10. The information retrieved by the resource broker 23 from the attribute repository 29 may be attributes such as CPU utilization, Random Access memory (RAM) capacity or Virtual Memory (VM) Capacity. The status of the resource nodes 30 may also be retrieved. A known way of implementing this step is by an information services component of Globus

Toolkit, the MDS (Monitoring and Discovery Service). The MDS can then be configured to gather a specific range of attributes and status of the resource nodes 30.

[0028] Next, the step of receiving 315 a first priority attribute and a second priority attribute for the matching of resource nodes 30 that may be best suitable for performing the job. These first and second priority attributes may be decided by a user based on the requirements of the job. For example, the job submitted may require very large RAM capacity for handling large amounts of variable data, the first priority attribute will then be RAM. The second priority may then be CPU utilization, depending on the user's requirements for the job.

[0029] The step of matching 320 available resource nodes 30 corresponding to the first and second priority attributes is then performed. The status of all resource nodes have already been checked 310 in earlier steps. Following which the step of recommending 325 the matching resource node 30 to the job submitter 25.

[0030] While the present description refers to a first and a second priority attribute, it will be understood by one skilled in the art that the present invention may not be limited to simply two priority attributes. While it is entirely possible to have several priority attributes for selection and predetermination by a user, it will be commonly accepted that there would be at least one priority attribute for matching resource nodes 30 and at least a second priority attribute to provide an alternative selection if more than one resource node 30 matches the first priority attribute.

[0031] Furthermore, the above description discloses a resource broker 23 recommending a single matching resource node 30 to the job submitter. However, having at least one recommendations or more recommendations is also entirely

possible. A separate algorithm or process may then further administer the at least one or more recommendations.

[0032] Examples of some possible scenarios during job brokering are highlighted below.

Scenario 1:

1st Priority: RAM, 2nd Priority: CPU Utilization
Resource Node A: 256 Mb RAM; 10% CPU Utilization
Resource Node B: 128 Mb RAM; 15% CPU Utilization
Resource Node C: 64 Mb RAM; 20% CPU Utilization

[0033] In scenario 1, the first priority attribute is based on RAM, that being the case, the Resource broker 23 will recommend the resource node A to the job submitter 25.

Scenario 2:

1st Priority: RAM, 2nd Priority: CPU Utilization
Resource Node A: 256 Mb RAM; 15% CPU Utilization
Resource Node B: 256 Mb RAM; 10% CPU Utilization
Resource Node C: 64 Mb RAM; 20% CPU Utilization

[0034] In scenario 2, the first priority attribute is also based on RAM and the second priority attribute is based on CPU utilization. The Resource broker 23 in this case will recommend the Resource Node B to the job submitter 25. The Resource Node A and B have the same RAM values, however based on the second priority attribute of CPU utilization, Resource Node B has less CPU utilization and therefore more free computing capacity. Therefore Resource Node B is recommended instead of Resource Node A.

[0035] Referring to FIG.4, the step of checking 225 the status and managing of the all current jobs starts with the step of polling 510 for the status of all current jobs running in the grid computing environment 10 by the job monitor 27. The job

monitor 27 is aware of the jobs submitted by the job submitter 25 to the resource nodes 30 and polls these jobs for their status. Examples of status of the jobs at the resource nodes 30 might be: "ACTIVE", "FAILED" or "DONE".

[0036] Upon completion of the polling step 510, any failed 515 job would result in the re-performing of job brokering excluding 530 the resource node 30 that had caused the failure. Following which the job would then be re-submitted 532 to a new recommended resource node 30 matched and recommended by the resource broker 23 based on the earlier first and second priority attributes.

[0037] If however, the number of auto re-submissions 520 exceed a predetermined (x) number of times, the method 200 would then proceed to inform user of job failure notification 525.

[0038] Informing user of job failure notification 525 may be by several methods and is also an option that may be predetermined by the user. Job failure notification 525 may be performed by sending email to the user, via Short Messaging Service (SMS) and by Paging Notification to the user.

[0039] If however, after polling 510, with no current jobs having failed, any completed 535 jobs returning a status of "DONE" would then further proceed to the step of completion notification 540. Informing user of job completion 540 is also an option that may be predetermined by the user. Job completion notification 540 may similarly be performed by sending email to the user, via Short Messaging Service (SMS) and by Paging Notification to the user.

[0040] All current jobs that have not failed 515 nor completed 535 will continue to be polled 510 by the job monitor 27 until the jobs either fail 515 and have been resubmitted 520 a predetermined number of times or are completed 535.

[0041] Referring to FIG.5, the step of performing job brokering 530 excluding failed resource node starts with the step of checking 610 the attribute repository 29 for attributes and status for all resource nodes. The resource node that failed during the previous immediate job submission is excluded 615.

[0042] The step of matching 620 available resource nodes 30 corresponding to the previously received 315 first and second priority attributes is then performed. Bearing in mind that the resource node 30 from the previously failed job submission has been excluded and the status of all resource nodes have been checked 610 again.

[0043] Following which the step of recommending 625 the matching resource node 30 to the job submitter 25 would then be performed allowing for the re-submission of the job.

[0044] It will be appreciated that various modifications and improvements can be made by a person skilled in the art without departing from the scope of the present invention.

Claims:

1. A method for monitoring and managing grid resources comprising of a plurality of resource nodes and at least one supervisor node in a grid computing environment, the method comprising:

   a) receiving one or more jobs;

   b) performing job brokering to match the one or more jobs to one or more resource nodes by a job broker of the supervisor node, wherein said job brokering comprises:

      b1) checking and retrieving attributes and status for all resource nodes from an attribute repository;

      b2) receiving a first priority attribute and a second priority attribute;

      b3) matching available resource nodes corresponding to the first and second priority attributes; and

      b4) recommending at least one matching resource from the available resource nodes;

   c) submitting the one or more jobs to the at least one recommended and matching resource node by a job submitter of the supervisor node; and

   d) checking status of and managing of all current jobs by a job monitor of the supervisor node, wherein the job monitor comprises:

      d1) polling status of all current jobs in the grid computing environment;

      d2) re-performing job brokering by excluding the resource node that caused any one of the current jobs to fail in the previous immediate job submission; and

      d3) re-submitting the current job/jobs to at least one new recommended and matching resource node provided that the current job has/jobs have not been re-submitted a predetermined number of times.

2. The method in accordance with claim 1, wherein the step d) further comprises:

d4) performing job failure notification if any one of the current jobs has failed and has been re-submitted a predetermined number of times.

3. The method in accordance with claim 1 or 2, wherein the step d) further comprises:

d5) performing job completion notification if any one of the current jobs has been completed.

4. The method in accordance with claim 2 or 3, wherein the job failure notification may be performed by any one of the following: email, SMS and paging notification.

5. The method in accordance with claim 3 or 4, wherein the job completion notification may be performed by any one of the following: email, SMS, and paging notification.

6. The method in accordance with any one of the preceding claims, wherein the plurality of resource nodes periodically sends attributes and status of the plurality of resource nodes to the attribute repository.

7. A system for monitoring and managing grid resources comprising of a plurality of resource nodes in a grid computing environment, the system comprising:

an attribute repository for storing attributes and status of the plurality of resource nodes; and

at least one supervisor node having:

a resource broker for retrieving attributes and status from the attribute repository, receiving at least a first and a second priority attribute, and matching the first and second priority attributes to the plurality of resource nodes and recommending at least one matching resource node to a job submitter; wherein the job submitter submits one or more jobs to the at least

one matching resource node in accordance with recommendation by the resource broker; and

a job monitor for polling status of all current jobs in the grid computing environment and re-submitting to the resource broker for re-brokering by excluding any one of the current jobs that failed in the previous immediate job submission, and re-submitting the current job/jobs to at least one new recommended and matching resource node provided that the current job has/jobs have not been re-submitted a predetermined number of times.

8. The system in accordance with claim 7, wherein the job monitor is adapted to perform job failure notification if a job has failed and has been re-submitted a predetermined number of times.

9. The system in accordance with claim 7 or 8, wherein the job monitor is adapted to perform job completion notification if a job has been completed.