

CLAIMS

What is claimed is:

1. A method for adjusting a performance of a computing system, comprising:
 - receiving by the computing system at least one command to execute a computational workload from a client;
 - allocating system resources of the computing system to the client by the computing system and executing the computational workload by the computing system;
 - applying by the computing system an identification method, the identification method identifying a first system resource of the system resources which limits an execution performance of the computational workload;
 - performing by the computing system a temporary adjustment of an allocation of the first system resource to the client;
 - evaluating by the computing system whether the temporary adjustment of the allocation of the first system resource to the client improves an execution performance of the computational workload;
 - if the temporary adjustment of the allocation of the first system resource to the client improves the execution performance of the computational workload, transmitting by the computing system to the client a message indicating that the allocation of the first system resource to the client can be permanently adjusted; and
 - in case of receiving by the computing system from the client an instruction to perform a permanent adjustment of the allocation of the first system resource to the client, performing by the computing system the permanent adjustment of the allocation of the first system resource to the client.
2. The method according to claim 1, the system resources being selected from the group comprising central processing unit (CPU) resources, graphic processing unit (GPU) resources, field programmable gate array (FPGA) resources, special purpose accelerator resources, network resources, random access memory resources, non-volatile data storage resources and database resources.

3. The method according to claim 1, the identification method comprising execution of at least one algorithm selected from the group comprising heuristic algorithms, machine learning algorithms, and algorithms for multivariate analysis.
4. The method according to claim 1, further comprising:
 - if the temporary adjustment of the allocation of the first system resource to the client does not improve the execution performance of the computational workload, adjusting the identification method;
 - applying by the computing system the adjusted identification method, the adjusted identification method identifying a second system resource of the system resources which limits an execution performance of the computational workload;
 - performing a temporary adjustment of the allocation of the second system resource to the client;
 - evaluating by the computing system whether the temporary adjustment of the allocation of the second system resource to the client improves the execution performance of the computational workload;
 - if the temporary adjustment of the allocation of the second system resource to the client improves the execution performance of the computational workload, transmitting by the computing system to the client a message indicating that the allocation of the second system resource to the client can be permanently adjusted; and
 - in case of receiving by the computing system from the client an instruction to perform a permanent adjustment of the allocation of the second system resource to the client, performing by the computing system the permanent adjustment of the allocation of the second system resource to the client.

5. The method according to claim 1, wherein identifying the first system resource comprises:
- classifying by the computing system a type of computation that is being performed or a type of software that is being run during execution of the computational workload; and
 - consulting by the computing system a resource usage profile associated with the type of computation or the type of software, the resource usage profile specifying that the first system resource is likely to limit the execution performance.
6. The method according to claim 1, wherein the temporary and/or permanent adjustment of the allocation of the first system resource to the client comprises increasing or decreasing a fraction of the first system resource allocated to the client.
7. The method according to claim 1, wherein evaluating by the computing system whether the temporary adjustment of the allocation of the first system resource to the client improves the execution performance of the computational workload comprises:
- quantifying a first execution performance of the computational workload before the temporary adjustment;
 - quantifying a second execution performance of the computational workload during the temporary adjustment; and
 - if the second execution performance surpasses the first execution performance, determining that the temporary adjustment improves the execution performance of the computational workload.
8. The method according to claim 7, further comprising:
- quantifying a third execution performance of the computational workload after the temporary adjustment of the allocation of the first system resource to the client; and
 - if the second execution performance surpasses the third execution performance in addition to surpassing the first execution performance, determining that the temporary adjustment improves the execution performance of the computational workload.

9. The method according to claim 1, further comprising:
including in the message to the client by the computing system data obtained by quantifying a first execution performance and a second execution performance, in order to provide the client with a prognosis regarding a performance enhancement which can be obtained by performing the permanent adjustment of the allocation of the first system resource to the client.
10. The method according to claim 9, further comprising:
calculating by the computing system a performance difference between the first execution performance and the second execution performance and including the performance difference in the message to the client.
11. The method according to claim 1, further comprising:
receiving by the computing system a first command from the client to execute a subtask of the computational workload; and
performing by the computing system the temporary adjustment of the allocation of the first system resource to the client while the subtask is being executed.
12. The method according to claim 1, further comprising:
receiving by the computing system a command from the client to execute a subtask of the computational workload and executing the subtask until completion;
after the subtask has been completed, performing by the computing system the temporary adjustment of the allocation of the first system resource to the client; and
re-executing the subtask of the computational workload with the temporary adjustment in effect.

13. The method according to claim 1, further comprising:
- receiving by the computing system a first command from the client to execute a first subtask of the computational workload and executing the first subtask until completion;
 - after the first subtask has been completed, performing by the computing system the temporary adjustment of the allocation of the first system resource to the client; and
 - receiving by the computing system a second command from the client to execute a second subtask of the computational workload and executing the second subtask of the computational workload.
14. The method according to claim 1, further comprising:
- inserting by the computing system into the message to the client a price for performing the permanent adjustment of the allocation of the first system resource to the client; and
 - in response to receiving the instruction from the client, if the instruction contains information specifying that the client agrees to the price, performing the permanent adjustment of the allocation of the first system resource to the client.
15. The method according to claim 1, the instruction from the client containing a bid and the method further comprising:
- determining by the computing system that a temporary adjustment of an allocation of the first system resource to further clients improves an execution performance of further computational workloads on the further clients;
 - transmitting by the computing system to the further clients, messages indicating that the allocation of the first system resource to the further clients can be permanently adjusted;
 - in response thereto, receiving by the computing system further instructions from the further clients to perform a permanent adjustment of the allocation of the first system resource to the further clients, the further instructions containing further bids; and
 - in response to receiving the instruction and the further instructions, if the bid is higher than the further bids, performing by the computing system the permanent adjustment of the allocation of the first system resource to the client.

16. The method according to claim 1, the computational workload being executed on a virtual machine of the computing system.
17. The method according to claim 1, the system resources of the computing system being distributed over multiple networked computing devices.
18. The method according to claim 1, wherein the computing system provides infrastructure as a service in a cloud computing environment by executing the computational workload.
19. A computing system, the computing system being configured to:
- receive at least one command to execute a computational workload from a client;
 - allocate system resources of the computing system to the client and to execute the computational workload;
 - apply an identification method, the identification method identifying a first system resource of the system resources which limits an execution performance of the computational workload;
 - perform a temporary adjustment of an allocation of the first system resource to the client;
 - evaluate whether the temporary adjustment of the allocation of the first system resource to the client improves an execution performance of the computational workload;
 - if the temporary adjustment of the allocation of the first system resource to the client improves the execution performance of the computational workload, transmit to the client a message indicating that the allocation of the first system resource to the client can be permanently adjusted; and
 - in case of receiving from the client an instruction to perform a permanent adjustment of the allocation of the first system resource to the client, perform the permanent adjustment of the allocation of the first system resource to the client.

20. A computer program product for operating a computing system, the computer program product comprising a computer readable storage medium having program instructions embodied therewith, the program instructions executable by a processor to cause the computing system to:

receive at least one command to execute a computational workload from a client;

allocate system resources of the computing system to the client and execute the computational workload;

apply an identification method, the identification method identifying a first system resource of the system resources which limits an execution performance of the computational workload;

perform a temporary adjustment of an allocation of the first system resource to the client;

evaluate whether the temporary adjustment of the allocation of the first system resource to the client improves an execution performance of the computational workload;

if the temporary adjustment of the allocation of the first system resource to the client improves the execution performance of the computational workload, transmit to the client a message indicating that the allocation of the first system resource to the client can be permanently adjusted; and

in case of receiving from the client an instruction to perform a permanent adjustment of the allocation of the first system resource to the client, perform the permanent adjustment of the allocation of the first system resource to the client.