

WHAT IS CLAIMED IS:

1. A computer-implemented method for dynamically managing requests for computing capacity provided by a program execution service, the method comprising:

under control of a program execution service that provides a plurality of computing nodes that are each configurable to execute one or more programs of multiple users of the program execution service,

receiving from a user of the program execution service a request to generate a virtualization environment for execution of a user application during a requested period of time, the request comprising information associated with the user application that permits the program execution service to execute a program based at least in part on the user application;

automatically generating the virtualization environment, the virtualization environment including one or more program services provided by the program execution service, the program services comprising: (1) a load balancer configured to distribute workload across one or more of the plurality of computing nodes or one or more instances of the virtualization environment, (2) a monitoring interface configured to allow the user to monitor execution of the program, (3) a load scaler configured to scale computing resources of the program execution service in response to variations in demand for execution of the program, and (4) a plurality of database management services;

determining from the plurality of computing nodes of the program execution service a group of computing nodes that can satisfy the request by the user;

executing one or more instances of the virtualization environment on the group of computing nodes during the requested period of time;

receiving from the user a request to make at least one change to the virtualization environment or to the one or more instances executing the virtualization environment on the group of computing nodes during the requested period of time; and

automatically implementing the at least one change during the requested period of time.

2. The method of Claim 1, further comprising providing the user with a computing interface for requesting generation of the virtualization environment or requesting changes to the virtualization environment.

3. The method of any one of the preceding claims, wherein the request to generate the virtualization environment further comprises a request for use by the program execution service of one or more of: (1) an operating system, (2) an application server, (3) one or more computing nodes having desired program execution capacity or geographic distribution, and (4) at least one of the plurality of database management services.

4. The method of any one of the preceding claims, further comprising:
automatically identifying failure of at least one of the one or more instances;
and

automatically launching at least one new instance for executing the virtualization environment on the group of computing nodes.

5. The method of any one of the preceding claims, further comprising:
automatically identifying failure of at least one of the group of computing nodes;

automatically allocating at least one new computing node from the plurality of computing nodes of the program execution service; and

automatically executing at least one new instance of the virtualization environment on the at least one new computing node.

6. A system configured to dynamically manage requests for computing capacity of a computing service, the system comprising:

a computer memory configured to store one or more program modules for dynamically managing requests for computing capacity of a computing service, the computing service comprising a plurality of computing nodes; and

an interactive request manager configured to communicate with the computer memory and to execute the one or more program modules stored in the computer memory, the program modules configured to:

receive a request from a user to generate a virtualization environment for execution of a user application during a usage period, the request

comprising information associated with the user application that permits the computing service to execute a program based at least in part on the user application;

generate the virtualization environment, the virtualization environment including one or more program services comprising (1) a load balancer configured to distribute workload across one or more of the plurality of computing nodes or one or more instances of the virtualization environment, (2) a monitoring interface configured to allow the user to monitor execution of the program, (3) a load scaler configured to scale computing resources in response to variations in demand for execution of the program, and (4) a plurality of database management services;

execute one or more instances of the virtualization environment on a group of one or more of the plurality of computing nodes during the usage period;

receive from the user a request to make at least one change to the virtualization environment or to the one or more instances executing the virtualization environment on the group of one or more of the plurality of computing nodes during the usage period; and

implement the at least one change during the usage period.

7. The system of Claim 6, wherein the plurality of computing nodes of the computing service comprises one or more physical computing systems.

8. The system of Claim 7, wherein the plurality of computing nodes comprises one or more virtual machines hosted on the one or more physical computing systems.

9. The system of any one of Claims 6-8, wherein the usage period comprises a start time when the execution of the one or more instances is to begin and an end time when execution of the one or more instances is to end.

10. The system of any one of Claims 6-9, wherein the program modules are further configured to determine from the plurality of computing nodes of the computing service a group of computing nodes that can satisfy the request by the user for generation of the virtualization environment.

11. The system of Claim 10, wherein the virtualization environment comprises the load balancer, and the load balancer is configured to balance execution of the one or more instances of the virtualization environment across the group of one or more of the plurality of computing nodes during the usage period.

12. The system of Claim 10, wherein program modules are further configured to:
identify failure of at least one of the one or more instances of the virtualization environment; and

launch at least one new instance of the virtualization environment on the group of one or more of the plurality of computing nodes.

13. The system of Claim 10, wherein the program modules are further configured to notify the user if an error is detected in the usage of the one or more instances of the virtualization environment or in the usage of the group of one or more of the plurality of computing nodes.

14. The system of Claim 10, wherein the request to make at least one change to the virtualization environment comprises a request to execute at least one instance of the virtualization environment on a computing node in a different geographic location than at least some of the computing nodes in the group of the one or more of the plurality of computing nodes.

15. The system of any one of Claims 6-14, wherein the request to generate the virtualization environment further comprises a request for execution of the user application in one or more desired geographic regions.

What is claimed is:

1. A computer-implemented method for dynamically managing requests for computing capacity provided by a program execution service (PES), the method comprising:

under control of the PES that provides a plurality of computing nodes that are each configurable to execute one or more programs of multiple users of the PES, the PES further providing a plurality of user-customizable program services, comprising (1) a user-customizable load balancer configured to allow a user to distribute workload across one or more of the plurality of computing nodes or one or more instances of the virtualization environment, (2) a monitoring interface configured to allow the user to monitor execution of the program, and (3) a load scaler configured to scale computing resources of the PES in response to variations in demand for execution of the program;

receiving from a user of the PES a request to generate a virtualization environment for execution of a user application during a requested period of time, the request comprising information associated with the user application that permits the PES to execute a program based at least in part on the user application;

automatically generating the virtualization environment based on the user application, the virtualization environment including a user-selectable operating system and at least one of the user-customizable program services provided by the PES;

determining from the plurality of computing nodes of the PES a group of computing nodes that can satisfy the request by the user;

executing one or more instances of the virtualization environment on the group of computing nodes during the requested period of time;

receiving from the user a request to make at least one change to the virtualization environment or to the one or more instances executing the virtualization environment on the group of computing nodes during the requested period of time, wherein the at least one change comprises a change to the user-selectable operating system; and

automatically implementing the at least one change during the requested period of time.

2. The method of claim 1, further comprising providing the user with a computing interface for requesting generation of the virtualization environment or requesting changes to the virtualization environment.

3. The method of claim 1, wherein the request to generate the virtualization environment further comprises a request for use by the PES of one or more of: (1) an application server, (2) one or more computing nodes having desired program execution capacity or geographic distribution, and (3) at least one of the plurality of database management services.

4. The method of claim 1, further comprising:
automatically identifying failure of at least one of the one or more instances; and
automatically launching at least one new instance for executing the virtualization environment on the group of computing nodes.

5. The method of claim 1, further comprising:
automatically identifying failure of at least one of the group of computing nodes;
automatically allocating at least one new computing node from the plurality of computing nodes of the PES; and
automatically executing at least one new instance of the virtualization environment on the at least one new computing node.

6. A system configured to dynamically manage requests for computing capacity of a computing service, the system comprising:

a computer memory configured to store one or more program modules for dynamically managing requests for computing capacity of a computing service, the computing service comprising a plurality of computing nodes that are each configured to execute one or more programs of multiple users of the PES; and

an interactive request manager configured to communicate with the computer memory and to execute the one or more program modules stored in the computer memory, the program modules configured to:

receive a request from a user to generate a virtualization environment from a plurality of virtualization environments available for selection by the user for execution of a user application during a usage period, the request comprising information associated with the user application that permits the computing service to execute a program based at least in part on the user application;
automatically generate the virtualization environment based on the user application, the virtualization environment including a user-selectable operating system and at

least one program service provided by the computing service, wherein the program service comprises: (1) a user-customizable load balancer configured to allow a user to distribute workload across one or more of the plurality of computing nodes or one or more instances of the virtualization environment, and one or more of: (2) a monitoring interface configured to allow the user to monitor execution of the program, (3) a load scaler configured to scale computing resources in response to variations in demand for execution of the program; execute one or more instances of the virtualization environment on a group of one or more of the plurality of computing nodes during the usage period; receive from the user a request to make at least one change to the virtualization environment or to the one or more instances executing the virtualization environment on the group of one or more of the plurality of computing nodes during the usage period, wherein the at least one change comprises a change to the user-selectable operating system; and automatically implement the at least one change during the usage period.

7. The system of claim 6, wherein the plurality of computing nodes of the computing service comprises one or more physical computing systems.

8. The system of claim 7, wherein the plurality of computing nodes comprises one or more virtual machines hosted on the one or more physical computing systems.

9. The system of claim 6, wherein the usage period comprises a start time when the execution of the one or more instances is to begin and an end time when execution of the one or more instances is to end.

10. The system of claim 6, wherein the request to generate the virtualization environment further comprises a request for use by the computing service of one or more of: (1) an application server, (2) one or more computing nodes having desired program execution capacity or geographic distribution, and (3) at least one of the plurality of database management services.

11. The system of claim 6, wherein the computing service provides an Application Programming Interface (API) for programmatically interacting with the computing service, and wherein the program modules are configured to receive the request from the user to generate or change the virtualization environment via the API.

12. The system of claim 6, wherein the program modules are further configured to determine from the plurality of computing nodes of the computing service a group of computing nodes that can satisfy the request by the user for generation of the virtualization environment.

13. The system of claim 6, wherein the virtualization environment comprises the monitoring interface, and the monitoring interface is configured to monitor usage of the group of one or more of the plurality of computing nodes during the usage period.

14. The system of claim 6, wherein the virtualization environment comprises the load balancer, and the load balancer is configured to balance execution of the one or more instances of the virtualization environment across the group of one or more of the plurality of computing nodes during the usage period.

15. The system of claim 12, wherein program modules are further configured to:

- identify failure of at least one of the one or more instances of the virtualization environment; and

- launch at least one new instance of the virtualization environment on the group of one or more of the plurality of computing nodes.

16. The system of claim 12, wherein program modules are further configured to:

- identify failure of at least one of the group of one or more of the plurality of computing nodes;

- allocate at least one new computing node from the plurality of computing nodes of the computing service; and

- execute at least one new instance of the virtualization environment on the at least one new computing node.

17. The system of claim 12, wherein the program modules are further configured to notify the user if an error is detected in the usage of the one or more instances of the virtualization environment or in the usage of the group of one or more of the plurality of computing nodes.

18. The system of claim 12, wherein the request to make at least one change to the virtualization environment comprises a request to execute at least one instance of the virtualization environment on a computing node in a different geographic location than at least some of the computing nodes in the group of the one or more of the plurality of computing nodes.

19. The system of claim 6, wherein the request to make at least one change to the virtualization environment comprises a request to terminate at least one of the one or more instances of the virtualization environment.

20. The system of claim 6, wherein the request to generate the virtualization environment further comprises a request for execution of the user application in one or more desired geographic regions.

21. A computing system configured to manage execution of programs for users, the system comprising:

- an interactive request manager component that is configured to manage execution of programs for users of a PES, the interactive request manager component configured to:

- receive from a user of the PES a request to generate a virtualization environment for execution of a user application during a usage period, the request comprising information associated with the user application that permits the PES to execute a program based at least in part on the user application, the PES providing a plurality of computing nodes configurable to execute programs of multiple users of the PES;

- automatically generate the virtualization environment based on the user application, the virtualization environment including a user-selectable operating system and at least one program service provided by the PES, wherein the program service comprises: (1) a user-customizable load balancer configured to allow a user to distribute workload across computing resources of the virtualization environment, and one or more of: (2) a monitoring interface configured to allow the user to monitor execution of the program, or (3) a load scaler configured to scale computing resources in response to variations in demand for execution of the program;

- execute one or more instances of the virtualization environment on a group of one or more computing nodes;

- determine at least one change to the virtualization environment or to the one or more instances executing the virtualization environment on the group of one or more of the computing nodes during execution of the one or more instances of the virtualization environment, wherein the at least one change comprises a change to the user-selectable operating system; and

automatically implement the at least one change during the execution of the one or more instances of the virtualization environment.

22. The system of claim 21, wherein the virtualization environment comprises an application container comprising an application stack configured to execute the user application.

23. The system of claim 21, wherein the virtualization environment is configured to permit the user application to spawn process threads or to open any port for communication with other applications.

24. The system of claim 23, wherein the port for communication with other applications comprises a hypertext transfer protocol secure port.

25. The system of claim 21, wherein the information associated with the user application comprises a web application archive file.

26. The system of claim 21, wherein the user application comprises a web application.

27. The system of claim 21, wherein the interactive request manager component is further configured to determine the at least one change to the virtualization environment or to the one or more instances by receiving a request from a second user to make the at least one change.

28. The system of claim 21, wherein the interactive request manager component is configured to determine the at least one change to the virtualization environment or to the one or more instances by monitoring execution of the instances.

29. The system of claim 21, wherein the at least one program service further comprises a plurality of database management services.