CLAIMS

1. A computer-implemented method for determining causality for dynamic modifications in program execution capacity provided to a user, the method comprising:

under control of one or more computing systems that are configured to provide a program execution service for use by multiple users, the program execution service having a plurality of computing nodes that are configurable to execute programs of the users of the program execution service,

receiving an indication of an initial desired program execution capacity for executing one or more software programs on behalf of a first user of the program execution service;

determining multiple capacity modification triggers for use in later initiating automated modifications to program execution capacity being provided to the first user, each of the capacity modification triggers including one or more criteria for use in determining if the capacity modification trigger is satisfied and including a specified type of program execution capacity modification to be requested if the capacity modification trigger is satisfied;

automatically determining a first group of multiple of the plurality of computing nodes for use in providing the initial desired program execution capacity of the first user, and making the computing nodes of the first group available at a first time to each execute one or more software programs for the first user;

automatically identifying multiple independent events that occur during a period of time between the first time and a later second time and that each are capable of resulting in a modification of the program execution capacity being provided by the multiple computing nodes of the first group, the multiple events including at least one of the specified capacity modification triggers being determined to be satisfied;

automatically determining an actual program execution capacity that is available to the first user from the computing nodes of the first group at the second

time, the actual program execution capacity at the second time being distinct from the desired program execution capacity provided to the first user at the first time based at least in part on one or more changes to availability of one or more computing nodes of the first group that occurred during a time period between the first time and the second time;

for each of the availability changes of the one or more computing nodes of the first group, automatically attributing one or more of the identified events as a cause of the availability change, at least one of the availability changes having a cause that is a combination of multiple events that are independent of each other and that include one or more of the at least one specified capacity modification triggers that are determined to be satisfied; and

providing an indication of the attributed cause for each of one or more of the availability changes so as to enable one or more further actions to be performed corresponding to that attributed cause.

2. The method of claim 1 wherein the multiple events further include at least one capacity modification instruction that is dynamically specified by the first user to request a specified modification to the program execution capacity being provided by the first group, and wherein the combination of multiple events that are attributed as the cause for one of the at least one availability changes includes one or more of the at least one specified capacity modification instructions.

3. The method of claim 1 wherein at least one other of the availability changes has an attributed cause that is a single one of the multiple events, the at least one other availability change being distinct from the at least one availability change.

4. The method of claim 3 further comprising automatically determining that the single one event is a determined type of cause of the at least one other availability change, the determined type of cause being either a direct cause or an

indirect cause, and providing an indication of the single one event being the determined type of cause of the at least one other availability change.

5. The method of the claim 1 wherein the automatic determining of the actual program execution capacity includes determining multiple availability changes that have occurred, and wherein the automatic attributing of the one or more events as the cause of each of the availability changes includes identifying one or more of the multiple events that are each a single cause of one of the multiple availability changes, and aggregating all of the multiple events that are not a single cause of the one of the multiple availability changes as part of the combination for each other of the multiple availability changes that does not have a single cause.

6. The method of claim 1 wherein, for one of the at least one availability changes, the automatic attributing of the combination of multiple events as the cause of the one availability change includes determining that the combination of multiple events is a determined type of cause of the one availability change, the determined type of cause being either a direct cause or an indirect cause, and wherein the providing of the indication of the attributed cause for each of the one or more availability changes includes providing an indication of the combination of multiple events being the determined type of cause of the one availability change.

7. The method of claim 1 wherein the multiple independent events each request a modification of the program execution capacity being provided by the first group, and wherein the method further comprises automatically determining a single aggregated modification to make to the program execution capacity of the first group based on aggregating the requested modifications of the multiple events, and initiating the determined single aggregated modification to the first group prior to the second time.

8. The method of claim 7 wherein the multiple events further include at least one capacity modification instruction that is dynamically specified by the first

user to request a specified modification to the program execution capacity being provided by the first group, wherein the determined multiple capacity modification triggers are specified by the first user prior to the first time, and wherein the combination of multiple events that are attributed as the cause for one of the at least one availability changes includes one or more of the at least one specified capacity modification instructions.

9. The method of claim 8 wherein the determined single aggregated modification is the one availability change for which the one or more specified capacity modification instructions are included in the combination that is attributed as the cause.

10. The method of claim 7 wherein the program execution capacity of the first group is measured based on a quantity of computing nodes that are part of the first group, such that the initial desired program execution capacity is an initial desired quantity of computing nodes and such that the determined actual program execution capacity at the second time is an actual quantity of computing nodes that are available as part of the first group at the second time, wherein the one or more availability changes of the one or more computing nodes of the first group each include at least one of one or more computing nodes that are part of the first group becoming unavailable and of one or more new computing nodes being added to the first group, and wherein the determined single aggregated modification to the first group is a determined change in quantity of the computing nodes for the first group.

11. The method of claim 10 wherein the determined single aggregated modification is to increase the quantity of computing nodes of the first group by a specified quantity of one or more computing nodes, wherein the determined single aggregated modification is one of the one or more availability changes for which the indication of the attributed cause is provided, and wherein the method further comprises performing one or more further actions that include initiating one or more

monetary charges to the first user based at least in part on the initiated determined single aggregated modification.

12. The method of claim 1 wherein one of the one or more availability changes for which the indication of the attributed cause is provided includes at least a first of the multiple events as at least part of the cause of the one availability change, and wherein the providing of the indication of the attributed cause for the one availability change is performed in response to a request for at least one of which events cause the one availability change and of which availability changes are caused by the first event.

13. The method of claim 12 wherein the request is received from a user, and wherein the providing of the indication of the attributed cause for the one availability change includes generating a human-readable explanation of the attributed cause for the one availability change and providing the generated human-readable explanation for display to the user.

14. The method of claim 1 further comprising, for each of one or more additional time periods that occur after the second time, automatically attributing one or more additional identified events as a cause of an availability change that occurred during the additional time period, and wherein, for one of the additional identified events that is the cause of one of the availability changes that occurred during one of the one or more additional time periods, the one additional identified event is a result of one of the availability changes that occurred during the time period between the first time and the second time.

15. The method of claim 1 wherein the making available of the computing nodes of the first group at the first time to each execute the one or more software programs for the first user includes automatically provisioning each of the computing nodes of the first group and includes automatically initiating execution of each of the one or more software programs on each of the computing nodes of the first group.

16. The method of claim 1 wherein the program execution service is a fee-based network-accessible service that is remote from the first user, wherein the program execution service provides at least one of one or more Application Programming Interfaces (APIs) for remote computing systems to programmatically interact with the program execution service over one or more networks and of a graphical user interface for use by remote users over the one or more networks, and wherein the first user pays one or more fees based at least in part on the attributed cause of one or more of the availability changes.

17. The method of claim 1 wherein the program execution service uses virtualization technology such that the plurality of computing nodes include, for each of multiple physical computing systems, multiple virtual machines hosted by the physical computing system that are each able to execute at least one program, wherein the computing nodes of the first group are hosted virtual machines, and wherein the one or more programs of the first user are part of a virtual machine image.

18. A computer-readable medium whose contents configure a computing system to determine causality for dynamic modifications in program execution capacity provided to a user, by performing a method comprising:

under control of the configured computing system,

receiving an indication of a first group of multiple computing nodes that are associated with a first user at a first time to provide an initial program execution capacity for use in executing one or more software programs on behalf of the first user;

automatically identifying multiple independent events that occur between the first time and a later second time and that each have an associated desired modification of the program execution capacity of the first group;

automatically determining an actual program execution capacity that is available to the first user from the computing nodes of the first group at the second time, the actual program execution capacity at the second time being distinct from the

initial program execution capacity provided at the first time based at least in part on one or more changes to availability of one or more computing nodes of the first group;

automatically attributing one or more causes to each of the determined availability changes of the first group, the one or more causes that are attributed to at least one of the availability changes being a combination of multiple of the identified independent events; and

providing an indication of the attributed one or more causes for one or more of the at least one availability changes.

19. The computer-readable medium of claim 18 wherein the configured computing system is part of a program execution service that provides a plurality of computing nodes configurable to execute programs of users of the program execution service, wherein the multiple computing nodes of the first group are a subset of the plurality of computing nodes, wherein the receiving of the indication of the first group includes receiving a request for the initial program execution capacity from the first user prior to the first time and automatically associating the first group of computing nodes with the first user at the first time for use in providing the initial program execution capacity to the first user, wherein the first user further indicates one or more capacity modification triggers prior to the first time for use in later initiating automated modifications to the program execution capacity of the first group, and wherein the multiple independent events include at least one of the capacity modification triggers being determined to be satisfied and include at least one capacity modification instruction that is dynamically specified by the first user.

20. The computer-readable medium of claim 19 wherein the desired modification of the program execution capacity for each of the multiple independent events is a requested modification, wherein the method further comprises automatically determining a single aggregated modification to make to increase the program execution capacity of the first group based on aggregating the requested modifications of the at least one satisfied capacity modification triggers and of the at

least one capacity modification instruction from the first user, and initiating the determined single aggregated modification prior to the second time, wherein the determined single aggregated modification is one of the at least one availability changes, and wherein the combination of the multiple independent events that is attributed as the one or more causes for the determined single aggregated modification includes the at least one satisfied capacity modification triggers and the at least one capacity modification instruction from the first user.

21. The computer-readable medium of claim 18 wherein the computer-readable medium is at least one of a memory of a computing system that stores the contents and a data transmission medium that includes a generated stored data signal containing the contents, and wherein the contents are instructions that when executed cause the computing system to perform the method.

22. A computing system configured to determine causality for dynamic modifications in program execution capacity for a user, comprising:

one or more processors; and

a system manager module that is configured to, when executed by at least one of the one or more processors, manage program execution capacity for multiple users of a network-accessible service by, for each of the multiple users:

determining an initial program execution capacity to be provided for use in executing one or more software programs on behalf of the user, and determining one or more capacity modification triggers for use in later initiating automated modifications to the program execution capacity being provided for the user;

automatically associating a group of multiple available computing nodes with the user at a first time for use in providing the initial desired program execution capacity to the user; and

after the first time,

automatically identifying multiple independent events that occur between the first time and a later second time and that each have an associated desired modification of the program execution capacity of the group, the multiple

events including at least one of the capacity modification triggers being determined to be satisfied;

automatically determining one or more changes to availability of one or more computing nodes of the group that affect the program execution capacity provided by the group;

automatically attributing one or more causes to each of the determined availability changes of the group, the one or more causes that are attributed to at least one of the availability changes being a combination of multiple of the identified independent events, and the one or more causes that are attributed to at least one other of the availability changes being a single one of the identified independent events; and

providing an indication of the attributed one or more causes for one or more of the determined availability changes of the group.

23. The computing system of claim 22 wherein the multiple independent events for one of the multiple users further include at least one capacity modification instruction that is dynamically specified by the one user, wherein the system manager module is further configured to automatically determine a single aggregated modification to make to increase the program execution capacity of the group for the one user based on aggregating the desired modifications of the at least one satisfied capacity modification triggers and of the at least one capacity modification instruction specified by the one user, and to initiate the determined single aggregated modification to the group for the one user prior to the second time, wherein the determined single aggregated modification is one of the at least one availability changes for the one user, and wherein the combination of the multiple independent events that is attributed as the one or more causes for the determined single aggregated modification includes the at least one satisfied capacity modification triggers and the at least one capacity modification instruction specified by the one user.

24. The computing system of claim 23 wherein the determined capacity modification triggers for the one user are specified by the one user, wherein the network-accessible service is a program execution service that provides a plurality of computing nodes configurable to execute programs of remote users of the program execution service, and wherein the multiple computing nodes of the group for each of at least one of the multiple users are a subset of the plurality of computing nodes.

25. The computing system of claim 22 wherein the system manager component includes software instructions for execution by the computing system.

26. The computing system of claim 22 wherein the system manager component consists of a means for managing program execution capacity for multiple users of the network-accessible service by, for each of the multiple users:

determining an initial program execution capacity to be provided for use in executing one or more software programs on behalf of the user, and determining one or more capacity modification triggers for use in later initiating automated modifications to the program execution capacity being provided for the user;

automatically associating a group of multiple available computing nodes with the user at a first time for use in providing the initial desired program execution capacity to the user; and

after the first time,

automatically identifying multiple independent events that occur between the first time and a later second time and that each have an associated desired modification of the program execution capacity of the group, the multiple events including at least one of the capacity modification triggers being determined to be satisfied;

automatically determining one or more changes to availability of one or more computing nodes of the group that affect the program execution capacity provided by the group;

automatically attributing one or more causes to each of the determined availability changes of the group, the one or more causes that are

attributed to at least one of the availability changes being a combination of multiple of the identified independent events, and the one or more causes that are attributed to at least one other of the availability changes being a single one of the identified independent events; and

providing an indication of the attributed one or more causes for one or more of the determined availability changes of the group.

CLAIMS

1. A computer-implemented method comprising:

providing, by one or more configured computing systems of a program execution service, a first group of multiple computing nodes for use in providing an initial program execution capacity at a first time for executing one or more software programs for a first user;

identifying, by the one or more configured computing systems, multiple events that occur between the first time and a later second time and that each are capable of resulting in a modification of the program execution capacity being provided by the multiple computing nodes of the first group;

determining, by the one or more configured computing systems, an actual program execution capacity that is available to the first user from the computing nodes of the first group at the second time, the actual program execution capacity at the second time being distinct from the initial program execution capacity provided at the first time based at least in part on a change to availability of one or more computing nodes of the first group that occurred between the first time and the second time;

attributing, by the one or more configured computing systems and at or after the second time, a combination of a plurality of the multiple identified events as a cause of the availability change; and

providing, by the one or more configured computing systems, an indication of the attributed cause for the availability change.

2. The method of claim 1 wherein the multiple events include at least one capacity modification request that is specified by the first user for a specified modification to the program execution capacity being provided by the first group, and wherein the combination of the multiple identified events that are attributed as the cause of the availability change includes one or more of the at least one specified capacity modification requests.

3. The method of claim 1 wherein the actual program execution capacity provided at the second time being distinct from the initial provided program execution capacity is further based in part on at least one other availability change that occurred between the first time and the second time, and wherein the method further comprises attributing a single one of the multiple identified events to the at least one other availability change.

4. The method of claim 3 further comprising determining that the single one event is a type of cause that is either a direct cause or an indirect cause, and providing an indication of the single one event being the determined type of cause.

5. The method of the claim 1 wherein the determining of the actual program execution capacity includes determining multiple other availability changes that have occurred, wherein the method further comprises attributing one or more of the identified events as the cause of each of the other availability changes, and wherein the attributing of the one or more identified events as the cause of each of the other availability changes includes identifying one or more of the multiple identified events that are each a single cause of one of the multiple other availability changes, and aggregating all of the multiple identified events that are not a single cause of the one of the multiple other availability changes as part of the combination for a first of the other multiple availability changes that does not have a single cause.

6. The method of claim 1 wherein the attributing of the combination of the plurality of identified events as the cause of the availability change includes determining that the combination is a determined type of cause of the availability change, the determined type of cause being either a direct cause or an indirect cause, and wherein the providing of the indication of the attributed cause for the availability change includes providing an indication of the combination being the determined type of cause of the availability change.

7. The method of claim 1 wherein the multiple events each indicate a modification of the program execution capacity being provided by the first group, and

wherein the method further comprises determining a single aggregated modification to the program execution capacity of the first group based on aggregating the indicated modifications of the multiple events, and initiating the determined single aggregated modification to the first group prior to the second time.

8. The method of claim 7 wherein the multiple events include at least one capacity modification request that is specified by the first user for a specified modification to the program execution capacity being provided by the first group, and wherein the combination of multiple events that are attributed as the cause for the availability change includes one or more of the at least one specified capacity modification requests.

9. The method of claim 7 wherein the program execution capacity of the first group is measured based on a quantity of computing nodes that are part of the first group, such that the initial program execution capacity is an initial quantity of computing nodes and such that the determined actual program execution capacity at the second time is an actual quantity of computing nodes that are available as part of the first group at the second time, wherein the availability change includes one or more computing nodes that are part of the first group becoming unavailable or one or more new computing nodes being added to the first group, and wherein the determined single aggregated modification to the first group is a determined change in quantity of the computing nodes for the first group.

10. The method of claim 9 wherein the determined single aggregated modification is to increase the quantity of computing nodes of the first group by a specified quantity of one or more computing nodes, wherein the determined single aggregated modification is the availability change for which the indication of the attributed cause is provided, and wherein the method further comprises performing one or more further actions that include initiating one or more monetary charges to the first user based at least in part on the initiated determined single aggregated modification.

11. The method of claim 1 wherein the availability change for which the indication of the attributed cause is provided includes at least a first of the multiple events as at least part of the cause of the one availability change, and wherein the providing of the indication of the attributed cause for the one availability change is performed in response to a request for an indication of which events cause the availability change.

12. The method of claim 11 wherein the request is received from a user, and wherein the providing of the indication of the attributed cause for the availability change includes generating a human-readable explanation of the attributed cause for the availability change and providing the generated human-readable explanation for display to the user.

13. The method of claim 1 further comprising, for each of one or more time periods that occur after the second time, automatically attributing one or more additional identified events as a cause of an additional availability change that occurred during the additional time period, and wherein, for one of the additional identified events that is the cause of one of the additional availability changes that occurred during one of the one or more time periods, the one additional identified event is a result of the availability change that occurred between the first time and the second time.

14. The method of claim 1 wherein the providing of the computing nodes of the first group at the first time for the executing of the one or more software programs for the first user includes provisioning each of the computing nodes of the first group and initiating execution of each of the one or more software programs on each of the computing nodes of the first group.

15. The method of claim 1 wherein the program execution service is a fee-based network-accessible service that is remote from the first user, wherein the program execution service provides at least one of one or more Application Programming Interfaces (APIs) for remote computing systems to programmatically interact with the program execution service over one or more networks and of a

graphical user interface for use by remote users over the one or more networks, and wherein the first user pays one or more fees based at least in part on the attributed cause of the availability change.

16. The method of claim 1 wherein the multiple computing nodes of the first group are part of a plurality of computing nodes that include, for each of multiple physical computing systems, multiple virtual machines hosted by the physical computing system that are each able to execute at least one program, wherein the computing nodes of the first group are hosted virtual machines, and wherein the one or more programs executed for the first user are part of a virtual machine image.

17. The computer-implemented method of claim 1 further comprising determining multiple capacity modification triggers for use in later initiating automated modifications to program execution capacity being provided to the first user, each of the capacity modification triggers including one or more criteria for use in determining if the capacity modification trigger is satisfied and including a specified type of program execution capacity modification to be requested if the capacity modification trigger is satisfied.

18. The computer-implemented method of claim 17 wherein the multiple identified events include at least one of the determined capacity modification triggers being determined to be satisfied, and wherein the combination of the multiple identified events that are attributed as the cause for the availability change includes one or more of the at least one determined capacity modification triggers that are determined to be satisfied.

19. The computer-implemented method of claim 18 wherein the multiple identified events further include at least one capacity modification instruction that is specified by the first user to request a specified modification to the program execution capacity being provided by the first group, and wherein the combination of the multiple identified events that are attributed as the cause for the availability change further includes one or more of the at least one capacity modification instructions.

20. A non-transitory computer-readable medium having stored contents that configure a computing system to:

receive an indication of a first group of multiple computing nodes that are associated with a first user at a first time to provide an initial program execution capacity for use in executing one or more software programs on behalf of the first user;

identify multiple events that occur between the first time and a later second time and that each have an associated desired modification of the program execution capacity of the first group;

determine an actual program execution capacity that is available to the first user from the computing nodes of the first group at the second time, the actual program execution capacity at the second time being distinct from the initial program execution capacity provided at the first time based at least in part on one or more changes to availability of one or more computing nodes of the first group;

attribute, at or after the second time, one or more causes to each of the determined availability changes of the first group, the one or more causes that are attributed to at least one of the availability changes being a combination of multiple of the identified events; and

provide an indication of the attributed one or more causes for one or more of the at least one availability changes.

21. The non-transitory computer-readable medium of claim 20 wherein the first user indicates one or more capacity modification triggers prior to the first time for use in later initiating automated modifications to the program execution capacity of the first group, and wherein the multiple events include at least one of the capacity modification triggers being determined to be satisfied.

22. The non-transitory computer-readable medium of claim 21 wherein the desired modification of the program execution capacity for each of the multiple events is a requested modification, wherein the stored contents further configure the computing system to determine a single aggregated modification to be made to increase the program execution capacity of the first group based on aggregating the requested

modifications of the at least one satisfied capacity modification triggers and of at least one capacity modification instruction from the first user, wherein the stored contents further configure the computing system to initiate the determined single aggregated modification prior to the second time, wherein the determined single aggregated modification is one of the at least one availability changes, and wherein the combination of the multiple events that is attributed as the one or more causes for the determined single aggregated modification includes the at least one satisfied capacity modification triggers and the at least one capacity modification instruction from the first user.

23. The non-transitory computer-readable medium of claim 20 wherein the multiple events include receiving an indication of at least one capacity modification instruction specified by the first user.

24. A computing system, comprising:

one or more processors; and

a system manager module that is configured to, when executed by at least one of the one or more processors:

associate a group of multiple available computing nodes with a user at a first time for use in providing an initial program execution capacity to the user; and

after the first time,

identify multiple events that occur between the first time and a later second time and that each have an associated indicated modification of the program execution capacity of the group;

determine one or more changes to availability of one or more computing nodes of the group that affect the program execution capacity provided by the group;

at or after the later second time, attribute one or more causes to each of the determined availability changes of the group, the one or more causes that are attributed to at least one of the availability changes being multiple of the identified events; and

provide an indication of the attributed one or more causes for one or more of the determined availability changes of the group.

25. The computing system of claim 24 wherein the multiple identified events further include at least one capacity modification instruction specified by the user, wherein the system manager module is further configured to determine a single aggregated modification to be made to increase the program execution capacity of the group based on aggregating the indicated modifications of the at least one satisfied capacity modification triggers and of the at least one capacity modification instruction specified by the one user, and to initiate the determined single aggregated modification to the group prior to the second time, wherein the determined single aggregated modification is one of the at least one availability changes, and wherein the multiple identified events that are attributed as the one or more causes for the determined single aggregated modification includes the at least one satisfied capacity modification triggers and the at least one capacity modification instruction specified by the user.

26. The computing system of claim 25 wherein the user is one of multiple users of the network-accessible service, wherein the determined capacity modification triggers are specified by the one user, wherein the network-accessible service is a program execution service that provides a plurality of computing nodes configurable to execute programs of the multiple users, and wherein the multiple computing nodes of the group are a subset of the plurality of computing nodes.

27. The computing system of claim 24 wherein the system manager module is further configured to determine one or more capacity modification triggers for use in later initiating automated modifications to the program execution capacity being provided for the user, and wherein the multiple identified events include at least one of the capacity modification triggers being determined to be satisfied.