CLAIMS

1. A system for profiling computing resource usage, the system comprising:

one or more processors;

a computer-readable memory including executable instructions that, when executed by the one or more processors, configure the system to:

calculate an expected operating constraint for an instance of a virtual machine based at least in part on operating metrics determined from running at least a prior instance of a similar virtual machine, wherein the expected operating constraint varies over a lifecycle of the instance of the virtual machine;

receive a request to instantiate the virtual machine; and in response to the request:

identify a host computing device, of a plurality of host computing devices, associated with one or more operating characteristics related to the operation of virtual machine instances based partly on whether the one or more characteristics satisfy the expected operating constraint over at least a portion of the lifecycle of the instance of the virtual machine; and

cause at least in part, a new instance of the virtual machine to be instantiated on the host computing device.

2. The system of claim 1, wherein at least one of the operating metrics relates to central processing unit (CPU) utilization, memory utilization, network utilization, hard disk utilization, or electrical power utilization.

3. The system of claim 1, wherein at least one of the one or more characteristics comprises memory capacity, central processing unit (CPU) capacity, network bandwidth, network latency, position within a network topology, instruction set, or variance of a performance metric.

4. The system of claim 1, wherein the operating constraint relates to an expected usage amount of a computing resource provided by the host computing device.

5. The system of claim 4, wherein the expected usage amount is further based at least in part on data received from a customer associated with the virtual machine.

6. The system of claim 1, wherein the module, when executed, is further configured to:

receive an additional operating metric regarding operation of the new instance on the host computing device, the additional operating metric related to the one or more operating characteristics; and

in response to determining, based on the additional operating metric, that the one or more operating characteristics no longer satisfy the operating constraint, transfer the new instance to a second computing device associated with one or more additional operating characteristics that satisfy the operating constraint.

7. A computer-implemented method for profiling computing resource usage, the computer-implemented method comprising:

receiving, by a data center management component comprising one or more computing devices, a request for initialization of a software workload associated with an operating profile, wherein the operating profile is based at least in part on a plurality of historical operating metrics associated with one or more lifecycles of the software workload, and wherein the historical operating metrics vary over the one or more lifecycles of the software workload; and

in response to the request:

calculating, based at least in part on the plurality of historical operating metrics, one or more expected operating characteristics associated with the software workload, wherein the one or more expected operating characteristics vary over the one or more lifecycles of the software workload;

identifying a computing device of a plurality of computing devices based at least in part on the operating profile and the one or more expected operating characteristics associated with the software workload; and

causing the software workload to be initialized on the computing device.

8. The computer-implemented method of claim 7, wherein the operating profile comprises a first expected resource usage amount associated with a first computing resource, wherein the first expected resource usage amount is based at least in part on a plurality of historical operating metrics regarding usage of the first resource, and wherein a characteristic of the one or more characteristic comprises availability of the first computing resource.

9. The computer-implemented method of claim 8, wherein each of the plurality of historical operating metrics regarding usage of the first resource is associated with a time that a measurement of usage was recorded, and wherein the operating profile is further based at least in part on the time that each of the plurality of historical operating metrics was

recorded.

10. The computer-implemented method of claim 7, further comprising determining the operating profile based at least in part on the plurality of historical operating metrics.

11. The computer-implemented method of claim 7, further comprising determining the operating profile based at least in part on a service level agreement with a customer associated with the software workload.

12. The computer-implemented method of claim 7, wherein the operating profile comprises a desired operating characteristic of the computing device, the method further comprising:

receiving a current operating metric regarding operation of the software workload on the computing device, the substantially current operating metric related to the one or more operating characteristics; and

in response to determining, based on the substantially current operating metric, that none of the one or more operating characteristics associated with the computing device correspond to the desired operating characteristic, transferring the software workload to a second computing device associated with an operating characteristic corresponding to the desired operating characteristic.

13. The computer-implemented method of claim 7, wherein the operating profile comprises a first predefined operating profile of a plurality of predefined operating profiles.

14. The computer-implemented method of claim 13, wherein the predefined operating profiles are associated with levels in an operating hierarchy, and wherein the first predefined operating profile is associated with higher level, in relation to the software workload, of the operating hierarchy.

15. The computer-implemented method of claim 7, wherein the operating profile comprises a customer-specific operating profile, and wherein the plurality of historical operating metrics are associated with initialization or use of the software workload by the customer.

## CLAIMS

WHAT IS CLAIMED IS:

1. A system for profiling computing resource usage, the system comprising:

one or more processors;

a computer-readable memory including executable instructions that, when executed by the one or more processors, configure the system to:

determine an operating constraint for an instance of a virtual machine based at least in part on operating metrics determined from running at least an instance of a similar virtual machine;

receive a request to instantiate the virtual machine; and

in response to the request:

identify a host computing device, of a plurality of host computing devices, associated with one or more operating characteristics related to the operation of virtual machine instances based partly on whether the one or more characteristics satisfy the operating constraint; and

cause at least in part, a new instance of the virtual machine to be instantiated on the host computing device.

2. The system of Claim 1, wherein at least one of the operating metrics relates to central processing unit (CPU) utilization, memory utilization, network utilization, hard disk utilization, or electrical power utilization.

3. The system of Claim 1, wherein at least one of the one or more characteristics comprises memory capacity, central processing unit (CPU) capacity, network bandwidth, network latency, position within a network topology, instruction set, or variance of a performance metric.

4. The system of Claim 1, wherein the operating constraint relates to an expected usage amount of a computing resource provided by the host computing device.

5. The system of Claim 4, wherein the expected usage amount is further based at least in part on data received from a customer associated with the virtual machine.

6. The system of Claim 1, wherein the module, when executed, is further configured to:

receive an additional operating metric regarding operation of the new instance on the host computing device, the additional operating metric related to the one or more operating characteristics; and

in response to determining, based on the additional operating metric, that the one or more operating characteristics no longer satisfy the operating constraint, transfer the new instance to a second computing device associated with one or more additional operating characteristics that satisfy the operating constraint.

7. A computer-implemented method for profiling computing resource usage, the computer-implemented method comprising:

receiving, by a data center management component comprising one or more computing devices, a request for initialization of a software workload associated with an operating profile, wherein the operating profile is based at least in part on a plurality of historical operating metrics associated with the software workload ; and

in response to the request:

identifying a computing device of a plurality of computing devices based at least in part on the operating profile and one or more operating characteristics associated with the computing device; and

causing the software workload to be initialized on the computing device.

8. The computer-implemented method of Claim 7, wherein the operating profile comprises a first expected resource usage amount associated with a first computing resource, wherein the first expected resource usage amount is based at least in part on a plurality of historical operating metrics regarding usage of the first resource, and wherein a characteristic of the one or more characteristic comprises availability of the first computing resource.

9. The computer-implemented method of Claim 8, wherein each of the plurality of historical operating metrics regarding usage of the first resource is associated with a time that a measurement of usage was recorded, and wherein the operating profile is further based at least in part on the time that each of the plurality of historical operating metrics was recorded.

10. The computer-implemented method of Claim 7, further comprising determining the operating profile based at least in part on the plurality of historical operating metrics.

11. The computer-implemented method of Claim 7, further comprising determining the operating profile based at least in part on a service level agreement with a customer associated with the software workload.

12. The computer-implemented method of Claim 11, further comprising:

receiving a substantially current operating metric regarding operation of the software workload on the computing device, the substantially current operating metric related to the one or more operating characteristics; and

in response to determining, based on the substantially current operating metric, that none of the one or more operating characteristics associated with the computing device correspond to the desired operating characteristic, transferring the software workload to a second computing device associated with an operating characteristic corresponding to the desired operating characteristic.

13. The computer-implemented method of Claim 7, wherein the operating profile comprises a first predefined operating profile of a plurality of predefined operating profiles.

14. The computer-implemented method of Claim 13, wherein the predefined operating profiles are associated with levels in an operating hierarchy, and wherein the first predefined operating profile is associated with higher level, in relation to the software workload, of the operating hierarchy.

15. The computer-implemented method of Claim 7, wherein the operating profile comprises a customer-specific operating profile, and wherein the plurality of historical operating metrics are associated with initialization or use of the software workload by the customer.