

CLAIMS

WHAT IS CLAIMED IS:

1. A system, comprising at least one computing device configured with processors and memory, the memory including instructions that upon execution cause the system to:
 - maintain an instance pool including a set of running instances that can be assigned to containers for performing tasks and then returned to the instance pool, each instance of the set of running instances having a corresponding instance type, wherein each of the instance types have a set of performance characteristics;
 - obtain a set of parameters for instantiating a container from at least information in a task definition, wherein the computing device is configured with one or more processors and memory to perform at least one of the following:
 - a) select an instance from the instance pool based on a corresponding instance type having a set of performance characteristics that satisfy a minimum number of parameters of the set of parameters for instantiating the container on an instance of the corresponding instance type, wherein the minimum number is specified in the task definition; and
 - b) select, upon determining that the set of parameters for instantiating the container specify an underprovisioned instance type, an instance of an improved instance type from the instance pool such that the improved instance type satisfies a second set of parameters, the second set of parameters including one or more parameters that exceed one or more parameters of the set of parameters; and
 - instantiate the container on the selected instance based at least in part on the task definition.
2. The system of claim 1, wherein the instructions that cause the system to select the instance from the instance pool further include instructions that, upon execution cause the system to:
 - analyze the set of parameters obtained from the task definition to generate a set of performance specifications for the instance, wherein the set of performance specifications for the instance includes at least one of: a CPU specification

for the instance, a memory specification for the instance, or a network bandwidth specification for the instance.

3. The system of claim 2, wherein the set of performance specifications for the instance includes a hardware requirement for the instance, the hardware requirement for the instance specifying at least one of: a field-programmable gate array requirement, an application-specific integrated circuit requirement, a hardware security module requirement, or a graphics processing unit requirement.

4. The system of claim 1, wherein the task definition describes a set of containers, each container of the set of containers corresponding to one or more container descriptions.

5. The system of claim 1, wherein the instructions that upon execution, further cause the system to

provision the instance by attaching a network interface to the instance to connect the instance to a virtual network of the customer and attach a temporary volume to the instance, the temporary volume usable by the container to read and write data while the container is running.

6. The system of claim 1, wherein an instance agent of the instance monitors the instance and provides one or more health metrics of the instance to a container management service provided by a computing resource service provider.

7. The system of claim 6, wherein the instance agent monitors the containers for one or more resource usage metrics, the one or more resource usage metrics usable to determine a resource cost associated with the container.

8. A method, comprising:
maintaining a pool of running instances that can be assigned to containers for performing tasks and then returned to the instance pool;
determining whether an instance type satisfies a set of parameters for instantiating a container on an instance of the instance type, the set of parameters for instantiating the container included in a container description of the container;

selecting an instance of the instance type from an instance pool further comprising one of selecting the instance, from the instance pool, with an instance type having a set of performance characteristics that satisfy a minimum number of parameters specified in the container description and selecting the instance, from the instance pool, with an improved instance type, the improved instance type determined such that the improved instance type satisfies a second set of parameters including one or more parameters that exceed one or more parameters of the set of parameters included in the container description; and

instantiating the container on the selected instance based on the container description.

9. The method of claim 8, wherein selecting the instance further comprises, upon determining there are no instances of the instance type in the instance pool, selecting the instance from the instance pool with an instance type having a set of performance characteristics that satisfy a minimum number of parameters of the set of parameters for instantiating the container on an instance of the instance type, wherein the minimum number of parameters is specified in the container description.

10. The method of claim 8, wherein the instance includes an instance agent that communicates with a container management service provided by a computing resource services provider the agent configured to apply security updates to the instance, the security updates received from the container management service.

11. The method of claim 8, wherein selecting an improved instance type is based at least in part on a determination that the set of parameters for instantiating the container specify an underprovisioned instance type

12. The method of claim 8, wherein provisioning the instance further comprises attaching a network interface to the instance, the network interface configured to allow communication with the container via a web services interface.

13. The method of claim 8, further comprising returning the instance to the instance pool as a result of the container being terminated.