ered. In one embodiment, the instructions are stored on a removable media device for reading by local or remote systems. In other embodiments, the logic or instructions are stored in a remote location for transfer through a computer network or over telephone lines. In yet other embodiments, the logic or instructions are stored within a given computer, CPU, GPU, or system.

A second action may be said to be "in response to" a first action independent of whether the second action results directly or indirectly from the first action. The second action may occur at a substantially later time than the first action and still be in response to the first action. Similarly, the second action may be said to be in response to the first action even if intervening actions take place between the first action and the second action, and even if one or more of the intervening actions directly cause the second action to be performed. For example, a second action may be in response to a first action if the first action sets a flag and a third action later initiates the second action whenever the flag is set.

To clarify the use of and to hereby provide notice to the public, the phrases "at least one of <A>, <B>, . . . and <N>" or "at least one of <A>, <B>, . . . <N>, or combinations thereof" or "<A>, <B>, . . . and/or <N>" are defined by the Applicant in the broadest sense, superseding any other implied definitions hereinbefore or hereinafter unless expressly asserted by the Applicant to the contrary, to mean one or more elements selected from the group comprising A, B, . . . and N. In other words, the phrases mean any combination of one or more of the elements A, B, . . . or N including any one element alone or the one element in combination with one or more of the other elements which may also include, in combination, additional elements not listed.

While various embodiments of the innovation have been described, it will be apparent to those of ordinary skill in the art that many more embodiments and implementations are possible within the scope of the innovation. Accordingly, the innovation is not to be restricted except in light of the attached claims and their equivalents.

What is claimed is:

1. An apparatus comprising:
a processor configured to:
create an indication of an allocation strategy, wherein the indication of the allocation strategy indicates the allocation strategy for allocation of external primary memory across one or more memory appliances, wherein the external primary memory is external to a client but is primary memory to the client, wherein the client is in communication with the apparatus or is included in the apparatus;
determine the allocation strategy in response to a request for external primary memory based on the indication of the allocation strategy; and
allocate one or more regions of the external primary memory across the one or more memory appliances in accordance with the allocation strategy,
wherein the external primary memory on the one or more memory appliances is accessible with client-side memory access independent of all central processing units of the one or more memory appliances, and
wherein a slab of the external primary memory is allocable at the client independent of the memory appliances, in response to a request at the client to allocate the slab, by a selection at the client of a subset of one of the one or more regions of the external primary

memory previously allocated and a mapping at the client of the slab of the external primary memory into a virtual address space.

2. The apparatus of claim 1, wherein the processor is configured to follow one or more steps and/or rules, which are included the indication of the allocation strategy, in a determination of how to allocate the external primary memory.

3. The apparatus of claim 1, wherein the indication of the allocation strategy includes a profile of a memory appliance, the client, and/or a user account.

4. The apparatus of claim 1, wherein the indication of the allocation strategy includes a priority setting associated with a user account.

5. The apparatus of claim 1, wherein the indication of the allocation strategy includes a time-of-day limit for a user account.

6. The apparatus of claim 1, wherein the indication of the allocation strategy includes a duration of use limit for a user account.

7. The apparatus of claim 1, wherein the indication of the allocation strategy includes a maximum external primary memory usage limit.

8. The apparatus of claim 1, wherein the indication of the allocation strategy includes a network bandwidth setting associated with the client and/or a user account that indicates the allocation strategy is to select the memory appliances having a corresponding network bandwidth.

9. The apparatus of claim 1, wherein the indication of the allocation strategy includes a network locality setting associated with the client and/or a user account that indicates the allocation strategy is to select any of the memory appliances that are within a threshold distance of the client that requested the external primary memory.

10. A method of provisioning external memory, the method comprising:
creating, by a configuration unit, an indication of an allocation strategy, the indication of the allocation strategy indicating the allocation strategy for allocation of an external primary memory;
providing the indication of the allocation strategy to an allocation unit; and
allocating, by the allocation unit, the external primary memory for a client based on the indication of the allocation strategy, wherein the external primary memory is external to the client but is primary memory to the client, wherein the external primary memory is accessible over a communication interface of the client with client-side memory access that is independent of a central processing unit of any memory appliance on which the external primary memory resides, and wherein a slab of the external primary memory is allocable at the client independent of any memory appliance on which the external primary memory resides, in response to a request at the client to allocate the slab, by the client selecting a subset of the external primary memory previously allocated and mapping the slab of the external primary memory into a virtual address space of the client.

11. The method of claim 10, wherein the client includes a virtual machine and the external primary memory is for use by the virtual machine.

12. The method of claim 11, wherein the virtual machine is a first virtual machine, the method further comprising migrating data stored in the external primary memory to a

second virtual machine by changing ownership of the external primary memory from the first virtual machine to the second virtual machine.

**13.** The method of claim **11**, further comprising resizing memory available to the virtual machine by allocating more or less external primary memory to the virtual machine.

**14.** The method of claim **10**, wherein the client includes a virtual machine on a first physical machine, the external primary memory is allocated to the virtual machine and includes data accessed by the virtual machine from the first physical machine, the method further comprising migrating the virtual machine from the first physical machine to a second physical machine without copying the data by accessing the data in the external primary memory by the virtual machine from the second physical machine.

**15.** The method of claim **10**, wherein the client includes a virtual machine on a first physical machine, the external primary memory is allocated to the virtual machine and includes data accessed by the virtual machine from the first physical machine, the method further comprising migrating the virtual machine from the first physical machine to a second physical machine without copying the data by discarding a cached copy of the data from the virtual machine.

**16.** A system comprising:

a communication interface configured to receive a request to allocate an external primary memory for a client, wherein the external primary memory is external to a client but is primary memory to the client; and

a processor configured to:

determine a number of regions to allocate across a plurality of memory appliances on a network as part of the external primary memory; and

send requests for allocation of the regions to the memory appliances over the network; wherein after the regions of the external primary memory are allocated on the memory appliances, the regions of the external primary memory are accessible with client-side memory access that is independent of central processing units of the memory appliances and wherein a slab of the external primary memory is allocable at the client independent of the memory appliances, in response to a request at the client to allocate the slab, by the client selecting a subset of one of the regions of the external primary memory previously allocated and mapping the slab of the external primary memory into a virtual address space of the client.

**17.** The system of claim **16** further comprising a memory, the memory comprising external memory allocation metadata descriptive of external memory allocations, the external memory allocations including the allocation of the regions distributed across the memory appliances on the network for the external primary memory.

**18.** The system of claim **17**, wherein the external memory allocation metadata is replicated to a management server on the network, and the external memory allocation metadata is recovered from the management server if the system fails.

**19.** The system of claim **16**, wherein the processor is further configured to:

determine, in response to a request to resize the external primary memory allocation, a second set of regions in which to divide the external primary memory; and

determine a distribution of the second set of regions across the memory appliances.

**20.** The system of claim **16**, wherein a distribution of the regions of the external primary memory is based on status information received from the memory appliances.

**21.** The system of claim **16**, wherein the processor is further configured to:

detect a second memory appliance on the network, the second memory appliance not included in the memory appliances across which the regions of the external primary memory are allocated; and

in response, determine availability of the second memory appliance for redistribution of at least one of the regions for the external primary memory.

**22.** A method comprising:

allocating an external primary memory for a client, wherein the external primary memory is external to the client but is primary memory to the client, wherein the external primary memory is accessible over a communication interface of the client with client-side memory access that is independent of a central processing unit of any memory appliance on which the external primary memory resides, and wherein a slab of the external primary memory is allocable at the client independently of any memory appliance on which the external primary memory resides, in response to a request at the client to allocate the slab, by the client selecting a subset of the external primary memory previously allocated and mapping the slab of the external primary memory into a virtual address space of the client, wherein the client includes a virtual machine on a first physical machine, the external primary memory is allocated to the virtual machine and includes data accessed by the virtual machine from the first physical machine; and

migrating the virtual machine from the first physical machine to a second physical machine without copying the data by accessing the data in the external primary memory by the virtual machine from the second physical machine.

* * * * *