or operations indicated as optional in the drawings may just be one example of components or operations that are optional. Components or operations not marked as optional may, in fact, be optional in some examples.

Furthermore, although specific components of innovations were described, methods, systems, and articles of manufacture consistent with the innovation may include additional or different components. For example, a processor may be implemented as a microprocessor, microcontroller, application specific integrated circuit (ASIC), discrete logic, or a combination of other type of circuits or logic. Similarly, memories may be DRAM, SRAM, Flash or any other type of memory. Flags, data, databases, tables, entities, and other data structures may be separately stored and managed, may be incorporated into a single memory or database, may be distributed, or may be logically and physically organized in many different ways. The components may operate independently or be part of a same program. The components may be resident on separate hardware, such as separate removable circuit boards, or share common hardware, such as a same memory and processor for implementing instructions from the memory. Programs may be parts of a single program, separate programs, or distributed across several memories and processors.

The respective logic, software or instructions for implementing the processes, methods and/or techniques discussed throughout this disclosure may be provided on computer-readable media or memories or other tangible media, such as a cache, buffer, RAM, removable media, hard drive, other computer readable storage media, or any other tangible media or any combination thereof. The tangible media include various types of volatile and nonvolatile storage media. The functions, acts or tasks illustrated in the figures or described herein may be executed in response to one or more sets of logic or instructions stored in or on computer readable media. The functions, acts or tasks are independent of the particular type of instructions set, storage media, processor or processing strategy and may be performed by software, hardware, integrated circuits, firmware, micro code, or any type of other processor, operating alone or in combination. Likewise, processing strategies may include multiprocessing, multitasking, parallel processing and/or any other processing strategy known now or later discovered. 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 local primary memory;

an interconnect; and

a processor configured to cause, in response to a memory allocation request from an application, allocation of a region of an external primary memory on a memory appliance, the external primary memory on the memory appliance accessible by the apparatus over the interconnect with client-side memory access, wherein the client-side memory access is independent of a central processing unit of the memory appliance, wherein the external primary memory is memory that is external to the apparatus but is primary memory to the apparatus, wherein the processor is further configured to cache data in the local primary memory that is accessed in the external primary memory on the memory appliance, wherein the memory allocation request is a first memory allocation request, and wherein the processor is further configured to:

allocate, in response to a second memory allocation request from the application, a slab of the external primary memory by:

a selection, at the apparatus and independent of the memory appliance, of a subset of the region of the external primary memory to be the slab, and

a mapping, at the apparatus, of the slab of the external primary memory to a virtual address space.

2. The apparatus of claim **1**, wherein the local primary memory is a higher numbered memory tier than at least one of an L1 cache, an L2 cache, or an L3 cache.

3. The apparatus of claim **1**, wherein the local primary memory is combined with the processor.

4. The apparatus of claim **1**, wherein the processor is further configured to migrate data that is stored in a portion of the region of the external primary memory from a first memory tier to a second memory tier.

5. The apparatus of claim **1**, wherein the processor is further configured to cause data for one or more portions of the region to be migrated to lower numbered tiers by a fault-in at a target memory tier.

6. The apparatus of claim **1**, wherein data in a portion of the region is read into the local primary memory in response to a request to read the data received from the application.

7. A system comprising:

a local primary memory;

an interconnect; and

a processor configured to cause, in response to a memory allocation request from an application, allocation of a region of an external primary memory included in a memory appliance, the external primary memory in the memory appliance accessible by the system over the

interconnect with client-side memory access, wherein the client-side memory access is independent of a central processing unit of the memory appliance, wherein the external primary memory is memory that is external to the system but is primary memory to the system, wherein the processor is further configured to operate the local primary memory as a cache for data accessed in the external primary memory included in the memory appliance, wherein the memory allocation request is a first memory allocation request, and wherein the processor is further configured to:

allocate, in response to a second memory allocation request from the application, a slab of the external primary memory by:

a selection, within the system and independent of the memory appliance, of a subset of the region of the external primary memory to be the slab, and

a mapping, within the system, of the slab of the external primary memory to a virtual address space.

8. The system of claim **7**, wherein the processor is further configured to cache, in the local primary memory, a portion of data stored in the slab of the external primary memory, the data accessed in a memory access operation performed on the slab of the external primary memory.

9. The system of claim **7**, wherein the client-side memory access includes Remote Direct Memory Access (RDMA).

10. The system of claim **7**, wherein the local primary memory includes a level 4 memory tier.

11. The system of claim **7**, wherein the processor is further configured to change data in a portion of the region from a first memory tier to a second memory tier.

12. The system of claim **7**, wherein the processor is further configured to cause data for one or more portions of the region of the external primary memory to be migrated to lower-numbered tiers in response to a fault incurred at a target memory tier.

13. The system of claim **7**, wherein data in a portion of the region is read into the external primary memory in response to a pre-fetch request.

14. A method comprising:

causing, in response to a first memory allocation request from an application at a client device, allocation of a region of an external primary memory that is included in a memory appliance, wherein the external primary memory is memory that is external to the client device but is primary memory to the client device;

allocating, in response to a second memory allocation request from the application, a slab of the external primary memory by selecting, independent of the memory appliance, a subset of the region of the external primary memory to be the slab, and mapping the slab of the external primary memory to a virtual address space at the client device;

accessing the external primary memory, which is included in the memory appliance, by the client device over an interconnect with client-side memory access, wherein the client-side memory access is independent of a central processing unit of the memory appliance; and

caching data read from the external primary memory in a local primary memory of the client device.

15. The method of claim **14**, wherein caching the data in the local primary memory includes providing a level 4 cache.

16. The method of claim **14** further comprising migrating data for one or more portions of the region of the external primary memory to lower-numbered tiers by faulting-in at a target memory tier.

17. The method of claim **14**, wherein the caching is in response to receiving, from the application, a request to read the data.

18. The method of claim **14**, wherein the caching is in response to a pre-fetch request.

19. A system comprising:

an interconnect; and

a processor configured to cause, in response to a first memory allocation request from an application, allocation of a region of an external primary memory included in a memory appliance, the external primary memory in the memory appliance accessible by the system over the interconnect with client-side memory access, wherein the client-side memory access is independent of a central processing unit of the memory appliance, wherein the external primary memory is memory that is external to the system but is primary memory to the system, wherein the processor is further configured to cache data in a processor cache that is read from the external primary memory included in the memory appliance, and wherein the processor is further configured to:

allocate, in response to a second memory allocation request from the application, a slab of the external primary memory by:

a selection, within the system and independent of the memory appliance, of a subset of the region of the external primary memory to be the slab, and

a mapping, within the system, of the slab of the external primary memory to a virtual address space.

* * * * *