

EFFICIENT LIVE-MIGRATION OF REMOTELY ACCESSED DATA

BACKGROUND

[001] Cloud computing refers to network-based computing in which collections of servers housed in data centers or “server farms” provide computational resources and data storage as needed to remote end users. Some cloud computing services provide access to software applications such as word processors and other commonly used applications to end users who interface with the applications through web browsers or other client-side software. Users’ electronic data files are usually stored in the server farm rather than on the users’ computing devices. Maintaining software applications and user data on a server farm simplifies management of end user computing devices. Some cloud computing services allow end users to execute software applications in virtual machines. In a public cloud computing environment, multiple users are able to launch virtual machines (VMs).

[002] Often times the data for a particular virtual machine is stored on one or more physical machines that are separate from the physical machine on which the virtual machine is instantiated. For a variety of reasons – load sharing, server maintenance, etc. – some or all of the data stored on a particular physical machine may be migrated to another physical machine. Live-migration of data is the process of moving data off of one physical machine to another physical machine while the virtual machine (or alternatively, a non-virtual processing entity) performs arbitrary reads and writes on the data.

SUMMARY

[003] This specification relates to live migration of data.

[004] This document describes a systematic method and system for moving data off of a storage computer to another storage computer while providing consistent and high performance access to the data to a third-party processing device (e.g., a virtual machine or an application) that is accessing the data remotely from another physical machine.

[004a] In a first aspect, the present disclosure provides a method performed by a data processing system, the method comprising: storing, in a first physical machine, a mapping for a data processing process that processes data stored in the first physical machine, the data

stored according to a plurality of data subsets; for each data subset, logically mapping in the mapping an address range for the data subset on the first physical machine to a respective logical address range for the data processing process; enabling read and write access to the data for the data processing process according to the logical addressing; and in response to a determination that a first data subset stored on the first physical machine is to be migrated to a second physical machine separate from the first physical machine, blocking read and write access for the data processing process to the first data subset while maintaining read and write access for the data processing process to the other data subsets.

[004b] In a second aspect, the present disclosure provides a system, comprising: a data processing device; and a non-transitory computer-readable memory in data communication with the data processing device and storing instructions executable by the data processing device and that upon such execution cause the data processing device to perform operations comprising: storing, in a first physical machine, a mapping for a data processing process that processes data stored in the first physical machine, the data stored according to a plurality of data subsets; for each data subset, logically mapping in the mapping an address range for the data subset on the first physical machine to a respective logical address range for the data processing process; enabling read and write access to the data for the data processing process according to the logical addressing; and in response to a determination that a first data subset stored on the first physical machine is to be migrated to a second physical machine separate from the first physical machine, blocking read and write access for the data processing process to the first data subset while maintaining read and write for the data processing process to the other data subsets.

[004c] In a third aspect, the present disclosure provides A non-transitory computer-readable memory in data communication with the data processing device and storing instructions executable by the data processing device and that upon such execution cause the data processing device to perform operations comprising: storing, in a first physical machine, a mapping for a data processing process that processes data stored in the first physical machine, the data stored according to a plurality of data subsets; for each data subset, logically mapping in the mapping an address range for the data subset on the first physical machine to a respective logical address range for the data processing process; enabling read and write access to the data for the data processing process according to the logical addressing; in response to a determination that a

first data subset stored on the first physical machine is to be migrated to a second physical machine separate from the first physical machine, blocking read and write access for the data processing process to the first data subset while maintaining read and write access for the data processing process to the other data subsets; and in response to a migration of the first data subset from the first physical machine to the second physical machine to store the data subset on the second physical machine, updating the mapping by logically mapping an address range for the first data subset on the second physical machine to the respective logical address range for the data processing process..

[005] In general, one innovative embodiment of the subject matter described in this specification can be embodied in methods that include the actions of storing, in a first physical machine, data for a data processing process running on a second physical machine

that is separate from the first physical machine, the storing data comprising storing the data according to a plurality of data subsets that are each exclusive of each other; for each data subset, logically mapping in a mapping, by the data processing process, an address range for the data subset on the first physical machine to a respective logical address range for the data processing process; enabling read and write access to the data by the data processing process according to the logical addressing; determining that a first data subset stored on the first physical machine is to be migrated to a third physical machine separate from the first physical machine and the second physical machine, and in response: blocking read and write access by the data processing process to the first data subset while maintaining read and write access by the data processing process to the other data subsets, migrating, from the first physical machine to the third physical machine, the first data subset to store the data subset on the third physical machine, and updating the mapping by logically mapping, by the data processing process, an address range for the first data subset on the third physical machine to the respective logical address range for the data processing process. Other embodiments of this aspect include corresponding systems, apparatus, and computer programs, configured to perform the actions of the methods, encoded on computer storage devices.

[006] In general, another embodiment of the subject matter described in this specification can be embodied in a system that includes a first physical machine storing data according to a plurality of data subsets that are each exclusive of each other; a second physical machine that is separate from the first physical machine and a virtual machine on the second physical machine having read and write access to the data stored on the first physical machine and that, for each data subset, logically maps, in a mapping, an address range for the data subset on the first physical machine to a respective logical address range for the virtual machine; wherein in response to a determination that a first data subset stored on the first physical machine is to be migrated to a third physical machine separate from the first physical machine and the second physical machine: the first data subset is migrated from the first physical machine to the third physical machine; read and write access to the first data subset for the virtual machine is blocked during the migration while read and write access by the virtual machine to the other data subsets is maintained; and the mapping is updated by the virtual machine to logically map an address range for the first data subset on the third physical machine to the respective logical address range for the virtual machine; and read and

write access to the first data subset for the virtual machine is restored after the migration of the first data subset from the first physical machine to the third physical machine. Other embodiments of this embodiment include corresponding methods, apparatus, and computer programs, configured to perform the actions of the methods, encoded on computer storage devices.

[007] Particular embodiments of the subject matter described in this specification can be implemented so as to realize one or more of the following advantages. The method provides better performance guarantees to the virtual machine than the other methods of live-migration such as pre-copy and post-copy live migration. For example, pre-copy live-migration considers the whole data address space as a whole migrating unit, and thus the entire address space exists completely on the originating side or completely on the destination side of the migration. To provide “live” access to the data while the migration is ongoing, the virtual machine is allowed to continue reading/writing the data, but writes are tracked so that changes can be resent to the receiving physical machine(s) of the live migration. This resending takes more read bandwidth, network bandwidth, CPU processing, and time. Furthermore, a busy virtual machine will typically be changing the data faster than the changes can be sent over the network to the destination; in such situations, the virtual machines access rate must be slowed, resulting in performance degradation.

[008] With post-copy live-migration, the virtual machine is informed of the destination physical machine and requests the destination physical machine for the data. The destination physical machine provides the data if the data is stored on the destination physical machine; otherwise the destination physical machine fetches the data from the originating physical machine which the data is being migrated and then provides the data. When the data must be fetched from the originating physical machine the virtual machine experiences an increased latency. With many accesses from the destination physical machine to the originating physical machine, there is a significant overall bandwidth performance degradation.

[009] The methods and systems described below, however, overcome some or all of these operational characteristics, resulting in an improvement in the technology area of data storage and management. By processing the migrating data in data subsets, which may be a chunk of X MB of data, or a page of data, etc., the migration is much more granular than pre-copy live-migration. The data subset undergoing migration is precluded from being accessed

CLAIMS

1. A method performed by a data processing system, the method comprising:
 - storing, in a first physical machine, a mapping for a data processing process that processes data stored in the first physical machine, the data stored according to a plurality of data subsets;
 - for each data subset, logically mapping in the mapping an address range for the data subset on the first physical machine to a respective logical address range for the data processing process;
 - enabling read and write access to the data for the data processing process according to the logical addressing; and
 - in response to a determination that a first data subset stored on the first physical machine is to be migrated to a second physical machine separate from the first physical machine, blocking read and write access for the data processing process to the first data subset while maintaining read and write access for the data processing process to the other data subsets.

2. The method of claim 1, wherein:
the determination that the first data subset stored on the first physical machine is to be migrated to the second physical machine is performed by the first physical machine; and
blocking read and write access for the data processing process to the first data subset while maintaining read and write access for the data processing process to the other data subsets comprises blocking, by the first physical machine, read and write access for the data processing process to the first data subset.
3. The method of claim 2, further comprising:
receiving, from the first physical machine by the data processing process, a notification of the migration of the first data subset to the second physical machine; and
wherein updating the mapping by logically mapping the address range for the first data subset on the second physical machine to the respective logical address range for the data processing process is done in response to the data processing process receiving the notification provided by the first physical machine.
4. The method of claim 3, wherein the notification instructs the data processing process to not access the first data subset.
5. The method of claim 3, further comprising:
receiving, from the second physical machine by the data processing process, a notification of the migration of the first data subset to the second physical machine; and
wherein updating the mapping by logically mapping the address range for the first data subset on the second physical machine to the respective logical address range for the data processing process is done in response to receiving the notification by the data processing process.

6. The method of claim 1, further comprising determining, by the data processing process, that the first data subset stored on the first physical machine is to be migrated to the second physical machine; and

migrating, from the first physical machine to the second physical machine, the first data subset to store the data subset on the second physical machine comprises:

instructing, by the data processing process, the first physical machine to send the first data subset to the data processing process; and

sending, by the data processing process, the first data subset received from the first physical machine to the second physical machine;

the updating the mapping is done by the data processing process.

7. The method of claim 1, further comprising restoring read and write access to the first data subset for the data processing process after the migration of the first data subset from the first physical machine to the second physical machine.

8. The method of claim 1, wherein the data subsets are data chunks logically mapped to a physical address range.

9. The method of claim 1, wherein the data subsets are data blocks.

10. A system, comprising:
a data processing device; and
a non-transitory computer-readable memory in data communication with the data processing device and storing instructions executable by the data processing device and that upon such execution cause the data processing device to perform operations comprising:
storing, in a first physical machine, a mapping for a data processing process that processes data stored in the first physical machine, the data stored according to a plurality of data subsets;
for each data subset, logically mapping in the mapping an address range for the data subset on the first physical machine to a respective logical address range for the data processing process;
enabling read and write access to the data for the data processing process according to the logical addressing; and
in response to a determination that a first data subset stored on the first physical machine is to be migrated to a second physical machine separate from the first physical machine,
blocking read and write access for the data processing process to the first data subset while maintaining read and write for the data processing process to the other data subsets.
11. The system of claim 10, wherein:
the determination that the first data subset stored on the first physical machine is to be migrated to the third physical machine is performed by the first physical machine; and
blocking read and write access for the data processing process to the first data subset while maintaining read and write access for the data processing process to the other data subsets comprises blocking, by the first physical machine, read and write access for the data processing process to the first data subset.
12. The method of claim 10, further comprising, in response to a migration of the first data subset from the first physical machine to the second physical machine to store the data subset on the second physical machine, updating the mapping by logically mapping an

address range for the first data subset on the second physical machine to the respective logical address range for the data processing process.

13. The system of claim 12, the operating further comprising:
receiving, from the first physical machine by the data processing process, a notification of the migration of the first data subset to the second physical machine; and
wherein updating the mapping by logically mapping the address range for the first data subset on the second physical machine to the respective logical address range for the data processing process is done in response to the data processing process receiving the notification provided by the first physical machine.
14. The system of claim 13, wherein the notification instructs the data processing process to not access the first data subset.
15. The system of claim 13, the operations further comprising:
receiving, from the second physical machine by the data processing process, a notification of the migration of the first data subset to the second physical machine; and
wherein updating the mapping by logically mapping the address range for the first data subset on the second physical machine to the respective logical address range for the data processing process is done in response to receiving the notification by the data processing process.
16. The system of claim 10, the operations further comprising determining, by the data processing process, that the first data subset stored on the first physical machine is to be migrated to the second physical machine ; and
migrating, from the first physical machine to the second physical machine, the first data subset to store the data subset on the second physical machine comprises:
instructing, by the data processing process, the first physical machine to send the first data subset to the data processing process; and
sending, by the data processing process, the first data subset received from the first physical machine to the second physical machine;
the updating the mapping is done by the data processing process.
17. The system of claim 10, wherein the data subsets are data chunks logically mapped to a physical address range.

18. The system of claim 10, wherein the data subsets are data blocks.

19. A non-transitory computer-readable memory in data communication with the data processing device and storing instructions executable by the data processing device and that upon such execution cause the data processing device to perform operations comprising:

storing, in a first physical machine, a mapping for a data processing process that processes data stored in the first physical machine, the data stored according to a plurality of data subsets;

for each data subset, logically mapping in the mapping an address range for the data subset on the first physical machine to a respective logical address range for the data processing process;

enabling read and write access to the data for the data processing process according to the logical addressing;

in response to a determination that a first data subset stored on the first physical machine is to be migrated to a second physical machine separate from the first physical machine, blocking read and write access for the data processing process to the first data subset while maintaining read and write access for the data processing process to the other data subsets; and in response to a migration of the first data subset from the first physical machine to the second physical machine to store the data subset on the second physical machine, updating the mapping by logically mapping an address range for the first data subset on the second physical machine to the respective logical address range for the data processing process.

20. The method of claim 1, further comprising, in response to a migration of the first data subset from the first physical machine to the second physical machine to store the data subset on the second physical machine, updating the mapping by logically mapping an address range for the first data subset on the second physical machine to the respective logical address range for the data processing process.