

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.

[005] In general, one innovative aspect 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, wherein the mapping is implemented using a watermark X such that all addresses below X are on the first physical machine, and all addresses at or above X are on the third physical machine. 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 aspect 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, wherein the mapping is implemented using a watermark X such that all addresses below X are on the first physical machine, and all addresses at or above X are on the third physical machine. Other embodiments of this aspect 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

by the virtual machine. Thus, tracking of writes need not be performed. Once a data subset is migrated, it does not need to be resent because all future accesses go directly to the destination side.

[0010] The virtual machine experiences virtually no impact to performance if it is not accessing the data subset being migrated. As the data subset size decreases, the amount of time the virtual machine must be blocked while waiting for the data subset to migrate decreases. Conversely, as the data subset size increases, the amount of metadata for storage mapping decreases. Thus, by selectively evaluating the trade-off of wait time v. mapping maintenance, a system administration may tailor data subset size for a particular application that results in an improved migration performance operation for the application.

[0011] While the migration techniques described below do utilize some bandwidth for overhead, the amount utilized is relatively small compared to the bandwidth utilized by pre-copy or post-copy migration. This is still yet another improvement to the technological field of data migration.

[0012] Because read and write access is blocked for the data subset undergoing migration, no overhead mechanism for tracking changes to the data subset is needed, nor is there a need to specifically order virtual machine accesses to the data subset for the purpose of migration. For example, if the data store is a disk and the virtual machine performs a write to a location while there is an outstanding read to the same location for the purpose of migration, then the result of the read access is undefined. The systems and methods herein preclude concurrent access to the same location by blocking the virtual machine from accessing the specific region being migrated.

[0013] The live migration described below may be stopped at any time without losing progress. This is in contrast to pre-copy live migration, which must completely start over if stopped, and also in contrast to post-copy live migration, which cannot be stopped after accesses for the virtual machine are switched to the destination side.

[0014] The details of one or more embodiments of the subject matter described in this specification are set forth in the accompanying drawings and the description below. Other features, aspects, and advantages of the subject matter will become apparent from the description, the drawings, and the claims.