Future Generation Computer Systems 50 (2015) 75–86

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Rethink the storage of virtual machine images in clouds

Xiaolin Xu, Hai Jin, Song Wu *, Yihong Wang

Services Computing Technology and System Lab, Huazhong University of Science and Technology, Wuhan, 430074, China
Cluster and Grid Computing Lab, Huazhong University of Science and Technology, Wuhan, 430074, China
School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, 430074, China

HIGHLIGHTS

- We propose a three-tier zone based VM image storage model.
- Zone-based model balances requirements of performance, resources, and overhead.
- The zones and caches in Zone-based model have high scalability and availability.
- A friendly VM placement strategy is proposed to improve IO performance of VMs.
- We evaluate some popular models and conclude their advantages and disadvantages.

ARTICLE INFO

Article history:
Received 15 May 2014
Received in revised form
30 September 2014
Accepted 1 October 2014
Available online 16 October 2014

Keywords:
Cloud computing
Virtual machine image
Zone
Image storage

ABSTRACT

As one of the most prevalent cloud services, Infrastructure-as-a-Service provides virtual machines (VMs) with high flexibility. How to effectively manage a huge amount of VM images becomes a big challenge. On one hand, images affect VMs' disk IO performance significantly, which is essential to the quality of services, especially for those having intensive disk IO workloads. On the other hand, they consume many storage resources and cause much management cost, which is cared by cloud managers. Current ways to optimize images usually focus on either improving performance or decreasing image size, which unfortunately cannot satisfy the requirements of high IO performance, low storage consumption, and low management cost simultaneously. Typically, high IO performance requires images storing close to VMs, but this increases redundant data and consumes extra storage at the same time. Besides, a closer image means more data stored in local disks rather than a normal shared storage, which increases management cost as well. In this paper, we analyze these requirements and potential tradeoffs among them, and propose Zone-based model to well balance the requirements. We partition computing nodes into many zones, and construct a shared storage in each zone to cache hot data for high IO performance and low storage consumption. In addition, we improve the normal Copy-on-Write and cache mechanisms, providing new image types and cache functions to enhance the eventual effectiveness. The evaluations show that, our solution improves IO performance by more than 100% in general and even 10 times while adopting a friendly VM placement strategy, and gets close or less storage consumption and management cost than the other models at the same time.

1. Introduction

Depending on the virtualization technology, clouds convert physical resources to virtual resources billed under the ''pay-as-you-go'' form, which significantly improves resource utilization and reduces users' costs. One of the basic and widely provided cloud services is Infrastructure-as-a-Service (IaaS) [1], which provides a customizable computing environment saved in a virtual machine (VM) image. Generally, a virtual machine runs on a physical machine, looking as a complete and independent operation environment, while an image acts as a virtual disk, storing the operating system and applications of a VM, just as the disk to a physical machine.

In IaaS clouds, most images are mounted as virtual disks to respond IO requests during VMs' running, and are stored as ordinary files to persist complete computing environment during VMs' shutting. With rapid growth of VMs in clouds, how to manage their images effectively becomes a big challenge. On one hand, storing and organizing of images affect VMs' IO performance. In a

* Corresponding author.
E-mail addresses: xxl@hust.edu.cn (X. Xu), hjin@hust.edu.cn (H. Jin), wusong@hust.edu.cn (S. Wu), yihwang@hust.edu.cn (Y. Wang).

http://dx.doi.org/10.1016/j.future.2014.10.004
0167-739X/

large-scale cloud, computing system and storage system are usually kept separate from each other. The visits to images may cause obvious network loads, which in turn lowers down VMs' performance. Although the Copy-on-Write (CoW) technology [2] allows VMs visiting local images rather than remote ones for reducing global network loads, it unfortunately increases the chances of hot spots appearing in base images at the same time [3]. On the other hand, management operations probably affect system status. Essentially, managing images is just managing files in the cloud. VM creation and migration equal to the distribution and migration of image files. These management operations occupy tremendous valuable network bandwidth. A larger image would cost more transferring time, and produce more management overhead. Meanwhile, it is unacceptable that lots of redundant image files waste many storage resources.

Therefore, from both users' and managers' perspectives, there are obvious requirements to manage images well, which include high IO performance, low management overhead, and high storage utilization. Unfortunately, traditional image storage systems cannot well meet these requirements simultaneously. Almost all the optimizations aim at improving IO performance, such as [4,5], while only a few of them consider the rest requirements, such as [6,7]. The main reason is that there are hard tradeoffs between IO performance and the other two requirements. On one hand, high IO performance and low management overhead are usually hard to meet at the same time. Precisely, a general way of improving IO performance is to cache image data at local disks as much as possible. However, lowering management overhead requires saving data in a remote central storage as much as possible. On the other hand, high IO performance and high storage utilization have conflicts as well. Increasing storage utilization needs reducing redundant data, and a feasible way is to save images in the central storage. However, this approach leads to a high sharing degree, in which many remote requests are produced and thus declines IO performance.

In this paper, we analyze above requirements and tradeoffs, and propose an optimized storage model trying to satisfy the requirements. In detail, first, we describe the background of image storage and format. Second, we analyze the requirements of managing images and the potential tradeoffs among them, and discuss some traditional image storage models. At the end, we propose a new image storage model, which can provide high IO performance without much extra management cost and storage consumption. Roughly speaking, we divide all computing nodes to many zones and construct a shared storage for each zone to cache hot data of images. Zone storage achieves the goal of sharing data among VMs. Not only does it avoid the bottleneck caused by global sharing, it also avoids the high storage consumption caused by pure local caches. The evaluations show that, our solution improves IO performance more obvious than the other optimized solutions with similar management cost and storage consumption. It improves IO performance by more than 100% in general and even 10 times while adopting the VM placement strategy that is friendly to our model. In addition, we also verify the scalability and availability of our model in multiple scenarios.

Our contributions are summarized as below:

- We analyze the requirements of managing VM images, and discuss why it is hard to meet them simultaneously in practice.
- We classify some popular image storage models in industry and academia, and discuss their advantages and disadvantages according to their characteristics and the requirements.
- We propose a zone-based image storage model with considering all requirements and present some key metrics to evaluate models deeply.

The rest of this paper is organized as follows: Section 2 describes the background and related work. We analyze requirements and tradeoffs and discuss traditional models in Section 3, propose our model in Section 4, and evaluate it in Section 5. Section 6 concludes this paper.

2. Background and related work

In this section, we first briefly discuss the background of VM images, including cloud architecture, image format, and image storage, and then we discuss the related works as well.

2.1. Typical cloud architecture for storing images

Generally, a typical cloud data center consists of management nodes, computing nodes, sharing storage, multiple switches and routers, and some other types of support equipment. Computing nodes provide core resources, such as CPUs, memory, and disks. Depending on these resources, we can create and run multiple VMs in the cloud. It is possible to store VM images in local disks of computing nodes. However, just a small amount of VMs with large-size images could exhaust the local disk of one node. So, a global sharing storage is provided to store images centrally. These computing nodes and the storage are connected through switches (Ethernet or Fiber). According to different image storage models, VMs have various ways to locate images. If the image of a VM is stored locally, the VM can locate data in the local disk without network transmission. However, if the image is stored in the global storage or in the disks of the other nodes, requests should be redirected and transferred through one or multiple switches.

2.2. Image format and image storage

Image format. Image format is a critical factor that affects images' management. The most original approach to create a VM is to generate a blank image file in a specified format as a virtual block device, based on which the operating system is installed as same as that on a physical machine. However, this process is unacceptable since it always spends tens of minutes. Fortunately, we can accelerate it by using advanced image formats with new features. This is why the approaches based on the Copy-on-Write mechanism [2] are widely adopted.

Different image formats have distinctive features and drawbacks. Take RAW and QCOW2 as examples, the former one has high compatibility and flexibility, while the latter one has optimized size and CoW ability. In our practice, the images in RAW format always act as base images to support the incremental images in QCOW2 format. In fact, allocating a large-size RAW image to each VM is unbearable in normal clouds. Leveraging the native support of CoW feature in QCOW2 format, multiple incremental images can share the same base image, which reduces redundant data greatly. However, too much sharing may transfer a large amount of requests to base images, occurring hot spots and resource contentions at the same time.

Image storage. The architecture of image storage has significant impacts on images' organization. Traditionally, there are two choices to store images, local solution and remote solution. The remote solution is usually adopted by large-scale data centers. These data centers are typically composed by multiple computing nodes, and all the nodes are connected to a global storage. In such clouds, VMs run in computing nodes while their images are saved in the shared storage. During VMs' running, images are remotely mounted as virtual disks through network. Depending on the shared storage, the remote solution supports many advanced features, such as live migration [8,9] and snapshotting. However, the separate placement of VMs and images requires high dependable network hardware to maintain the performance of virtual disks, which unfortunately is incompatible with commercial equipment in general clouds.