Computer Science Laboratory Environment Using Docker

by Justinus Andjarwirawan

Computer Science Laboratory Environment Using Docker

Justinus Andjarwirawan[1,*], *Henry* Novianus Palit[1] , and *Adi* Kurniawan[1]

[1]Informatics Engineering, Petra Christian University, Siwalankerto 121-131, Surabaya, Indonesia

Abstract. Docker is an application that is based on open source technology that allows developers or anyone to create, run, do experiments and launch application in a container. Docker make the process of packaging application components together quickly in an insolated container, so it can run in local infrastructure without changing configuration on the container. Docker is also very light and fast when compared to hypervisor-based virtual machine. A university has a considerable number of students, while the number of computers in the department is not as many as the students. It makes the students not comfortable working on one computer, because configuring a computer on every lesson is different. Therefore, implementation of Docker application carried out in the laboratory. Docker and RAID storage system were implemented, testing was done by measuring the speed of Docker image transfer, and its effectiveness for laboratory programs.

Keywords: Docker, container, virtual machine, RAID, storage, computer lab, lab setup.

1 Introduction

Docker is an open source platform for developers, which is useful for building, distributing and running applications anywhere. Docker makes the process of packaging applications together with their components, dependencies and libraries quickly in an isolated container, so that they can be run in local infrastructure without further configuration or configuration in the container. Docker is also very lightweight and fast when compared to hypervisor-based virtual machines.

In a university computer lab, it is very common to share computers. By sharing this computer, a student cannot safely store data on that computer, because other people can read, edit and delete.

A computer laboratory, which every semester changes its program, each computer is reinstalled again to clean user data. It takes a long time to reinstall one computer at a time and install the applications and their dependencies.

Seeing the ability of this lightweight Docker, Docker has the potential to help improve data security that is owned by students and simplify the work of lab assistants. By providing virtual machines for each student, each student can conveniently store their data, and for lab assistants they do not have to reinstall each computer in the lab every semester.

[*] Corresponding author: justin@petra.ac.id

Another study of Docker has implemented a performance analysis comparing container technology to a traditional virtualization [6][7] and the result showed that there was a significant performance gain on the boot process when using Docker as the operating system used in Docker did not have as many startup executions as those in non-containerized environment.

Docker can also serve a service in a distributed platform [5] such as database, where CPU utilization can be distributed to perform concurrent requests.

A related work to this research is a Docker-based framework for Cybersecurity Labs [8] where some cases need a lot of effort in configuring the lab environment like the network configurations, therefore the lab materials are presented in a containers instead.

2 Theory

Docker is an open platform for anyone who aims to use a platform to build, distribute and run applications anywhere such as laptops, data centers, virtual machines or cloud. [1]

Docker uses a client-server architecture. Docker client contacts Docker daemon, which does the hard work, runs, and distributes the Docker container. Both Docker client and daemon can run on the same system. The Docker client and the daemon communicate via sockets or through the API provided by Docker. [2]

A virtual machine consists of a set of specifications and configuration files and is supported by the physical resources of a host. Each virtual machine has a virtual device that provides the same functionality as physical hardware and has additional benefits in terms of portability, management and security. [3]

The difference between Docker and virtual machines is that containers in Docker have the same isolated power and benefit of allocations such as virtual machines but different architectural differences allow containers to be more portable and efficient. In that case Docker can be operated on both virtual machines and non-virtual machines (bare metal).

RAID is a storage system that combines several hard disks and creates a large virtual hard disk. Over time, RAID continues to grow, has many new solutions to old problems, such as disk performance, redundancy, and scalability. With RAID, we can also make sure when there is a faulty hard drive, the server will still continue to work. The popular setup of RAID is mirror, the type of RAID that uses 2 or more hard disks that contain the same copy. If one hard drive has a problem, then the mirror hard disk can take over the task. [4]

3 System design

The design will involve a computer lab setup with a Local Area Network switch, computers and storage configuration in every computer in the lab as it is prepared for the new environment with Docker.

3.1 Network design

The network topology used is the star topology. This topology basically utilizes one of the important hardware in the formation of computer networks, namely the hub or switch. The advantage of using this topology is that when there is one faulty computer, it will not affect other computer in the network. Detecting problems will be easier when there is something wrong with the network. The disadvantage of this topology is that when the switch is broken, all computers cannot connect each other let alone the Internet, and also star topology will require more LAN cables.

3.1 Storage design

This project will use 3 forms of storage systems. The first is LVM (Logical Volume Manager), this storage system is provided automatically from Ubuntu itself. The way LVM works is by combining two pieces of partition from a different hard drive into one logical volume, but filling in the data is by populating one hard drive first, then filling the second hard drive. RAID storge system is a feature that is provided for most servers. RAID can be implemented by hardware-based when the hardware is available, or by software-based where Linux always provides. The first common software-based RAID is RAID 1. This RAID 1 is a RAID that works by mirroring. This RAID working system is combining 2 different partitions of the hard drive into 1 logical partition. The use of RAID 1 is when data is entered into a partition, the data will be written on two different hard drives.

The second RAID is RAID 0, which is striping. This RAID works by combining 2 different hard disks into 1 logical volume by adding them. The system works almost the same as LVM, but in RAID 0, the data is fragmented and inserted alternately to the first hard drive then to the second hard drive.

RAID is setup at the beginning before the LVM configuration. It must be determined which RAID configuration to use before everything else is set. RAID 1 will be the best solution for a high availability system where there is always at least a backup. RAID 0 is best when the storage is in a mechanical drive that needs capacity and read write performance but with no redundancy. LVM will follow the given partition size coming from RAID setup and it is scalable when there is another set of RAID system.

3.2 Docker

Docker is an open source software that can run on Linux and also on Windows' Microsoft Hyper-V hypervisor. The benefit of using Docker is that Docker is lightweight. Containers run on one machine that shares the same kernel operating system, so it can start directly, and make RAM use more efficient. Images are built on the file system layer so that they can share common files, make disk usage and download images much more efficient.

Containers isolate applications from each other and basic infrastructure while providing an additional layer of protection for applications.

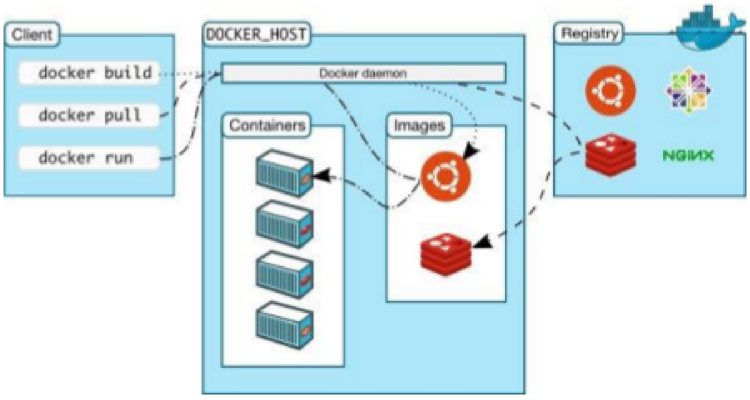

Fig. 1. Docker architecture

(Source: https://docs.docker.com/engine/docker-overview/)

Because it is built on facilities provided by the Linux Kernel, a Docker Container, unlike a virtual machine, does not require or include a separate operating system. Instead, it depends on kernel functions and uses resource isolation (CPU, memory, I/O blocks, network, etc.)

and a separate namespace to isolate the appearance of applications from the operating system, as seen in Figure 1. Docker accesses the virtualization feature the Linux kernel uses either the libcontainer library, which is available as Docker 0.9, or indirectly through libvirt, LXC (Linux Containers) or systemd-nspawn.

The registry docker is a centralized repository of a collection of Docker image, both public and private repository. The Docker public registry is called Docker Hub. Here one can push their own image or pull an image.

4 System Implementation

The implementation covers server and client configuration as well as the network in the lab. The server will store container data so that clients can access them remotely. A similar environment has also been implemented [9] to create a cloud type of service.

4.1 Server setup

First the server needs to install the Private Docker Registry. It will need a certificate using openssl. After running the command to create the certificate, entry to the CN section (Common Name) uses the domain name that will be used, as this project uses the domain name "registry.infor".

Because the default registry runs on port 5000 and each image push or pull uses port 5000, using Apache web server, the port is changed from 5000 to port 443, the https port, so there is no need to mention the port again when doing push or pull images as it is using the default port for an https connection.

Apache authentication is set for the clients to connect to this server. The authentication is using the http authentication module.

4.2 Client setup

This client computer will later be used by students to work in their daily activities in the laboratory. This computer is used only for editing images, but not as a storage area for images.

The first thing to do to configure this client computer based on Ubuntu Server is to install the docker program first. The step of installing a docker-engine for an Ubuntu Server-based client computer is exactly the same as the docker installation step on the server computer.

In Linux based client, after the installation of the docker-engine application is complete, the next thing to do is to add the private registry configuration. At this point the laboratorium is prepared to install and run selected applications.

5 Evaluation

This test will begin by running a container. In this test, several applications are installed in the container and that will be sent to the private registry. After the application has been installed, the container will return again to become a new image.

There is already a previous base image on the server, so that when a user enters the image that was created earlier based on the existing Ubuntu image, the image sent is only the top layer image, we can see at the top layer only the words "Pushed", while the other is "Layer already exists" which states that the server already has that data.

When the user pulls the image back, the docker only pulls the last layer part. This is because the client computer still has an Ubuntu image whose layer is the same, so the "Already exists" statement appears.

This shows that the Docker application is very light. Docker only wraps the applications that are installed, the Docker also wraps the configuration in a container that is very light because it does not need a base image.

5.1 RAID system storage transfer performance evaluation

After seeing the results of image transfer in Table 1 and 2, according to different storage systems, RAID 1 takes more time, because this storage system reads from 2 hard disks with the same file, but this RAID is much safer, because if 1 hard drive is faulty, then the other hard drive can run alone, because the data is the same.

The fastest transfer is to use RAID 0. Faster speed because the hard disk reads different files on each hard disk, this makes the hard disk work to transfer images faster because each hard disk sends different pieces of image, but there is no backup, if one drive fails, the whole data is gone because partial data from one drive cannot represent the whole data.

As a performance test purpose, RAID 0 and RAID 1 are performed as they represents two common base technology of RAID.

Table 1. RAID 1 data transfer performance

Number of PCs	Number of tests	Average time
1	5	0:09:375
2	5	0:14:005
4	5	0:51:232
6	5	1:36:493
8	5	1:11:124

Tabel 2. RAID 0 data transfer performance

Number of PCs	Number of tests	Average time
1	5	0:11:306
2	5	0:13:617
4	5	0:21:325
6	5	0:29:739
8	5	0:40:358

The test was done on Ubuntu Server 14.04 at the time of evaluation. Running on Core i5 PCs with 8GB of RAM and several 1TB hard drives.

Conclusion

Based on the results of the test it can be concluded several things as follows:

The applications and their dependencies are conveniently available to different lab sessions and the switching time between containers does not suffer the lab schedule. This is the most contributed advantage to using Docker.

Evaluation of RAID storage systems alone can accelerate the transfer speed of an image, so that the use of a hard drive can be fully utilized.

On the other hand, the implementation of this docker application also has drawbacks, namely the container only runs fully on the Ubuntu Linux operating system, some operating system functions are only available when users take full control of the operating systems. For that the students need to have their own virtual machines, also when using Microsoft Windows applications.

Using flash-based storage or SSD will significantly upgrade the read write performance and RAID as the CPU utilization and memory are low.

References

1. Docker Overview. [Online] from https://docs.docker.com/engine/docker-overview/ [Accessed on 27 July 2019].
2. Docker Docs. [Online] https://hub.docker.com/_/docker-docs [Accessed on 27 July 2019].
3. What Is a Virtual Machine? [Online] from https://pubs.vmware.com/vsphere-50/index.jsp?topic=%2Fcom.vmware.vsphere.vm_admin.doc_50%2FGUID-CEFF6D89-8C19-4143-8C26-4B6D6734D2CB.html [Accessed on 27 July 2019].
4. Vadala, Derek. *Managing RAID on LINUX: Fast, Scalable, Reliable Data Storage*. O'Reilly Media (2002).
5. Liu Lijuan. *Research and Implementation of Docker Performance Service in Distributed Platform*. International Journal of Engineering and Computer Science ISSN: 2319-7242. Vol. 6 Issue 11 (2017).
6. B.B. Rad, H.J. Bhatti, M. Ahmadi. *An Introduction to Docker and Analysis of its Performance*. IJCSNS International Journal of Computer Science and Network Security, Vol. 17 No. 3 (2017).
7. Carl Boettiger. *An Introduction to Docker for Reproducible Research, with Examples from the R Environment*. ACM SIGOPS Operating Systems Review, arXiv: 1410.0846v1 (2014).
8. C.E. Irvine, M.F. Thompson, M. McCarrin, J. Khosalim. Labtainers: *A Docker-based Framework for Cybersecurity Labs*. USENIX Workshop on Advances in Security Education, Vancouver, B.C., Canada (2017).
9. John Fink. Docker: *A Software as a Service, Operating System-Level Virtualization Framework*. Code4Lib Journal Issue 25 (2014).

Computer Science Laboratory Environment Using Docker

ORIGINALITY REPORT

0%	0%	0%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		