Large-scale analysis of Docker registry dataset

Your N. Here
Your Institution

Second Name
Second Institution

Abstract

Docker containers are becoming increasingly popular due to their isolation properties, low overhead, and efficient packaging of execution environment. Docker containers are created from Docker images which package software dependencies, environment configuration, and other parameters that affect application's runtime. Docker registry stores images and allows clients to push images to and pull images from it. The total amount of images stored in Docker Hub registry is 457,627 now and is continuing to increase. As the amount of images stored in Docker registry increases it becomes extremely important to understand images' statistic characteristics for registry designer to deploy and optimize a registry and for docker client to design a container.

Our goal is to collect statistics from a large amount of Docker images and perform a large-scale characterization of Docker images. So far, we downloaded 143,784 images from Docker Hub registry (20TB, 31% of the whole dataset) and performed the first in-depth analysis for these images. We characterize images using multiple metrics, e.g., image size distribution, file size and type distribution, the number of layers per image, the amount of redundant data between images and layers, temporal trends.

1 Introduction

outline:
1. container and registry are becoming popular
2. but we dont know the statistic characterization about registry
3. This paper provide a first in-depth analysis
4. contribution: interesting findings

2 Background

3 Methodology

outline:
1. our method to collect and download images (how, pull, extracting, manifest+layer tarball, server setup)
2. our Dataset (size, table)

4 Results

4.1 Image

4.2 Layer

5 Relatedwork

do we need a relatedwork?

6 Conclusion