Simple Object Storage - Report 2

Nijad Huseynov

June 2023

As part of the research class, I am implementing a distributed object storage for small files. When the system is ready, I will upload the collected data to the system, validate its correctness, and measure its performance for read and write under different workloads. Based on the results, I may change or improve some parts of the system. So this requires a quantitative research approach to evaluate the system. Figure 1 describes planned research steps.

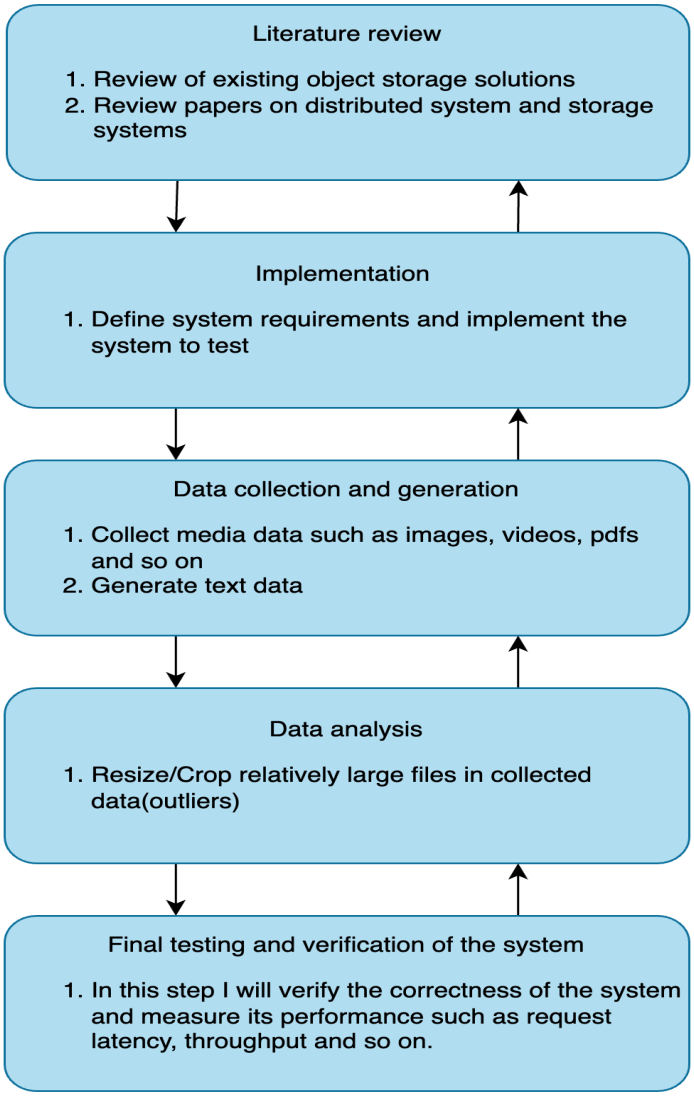

Figure 1: Flow diagram of the research

This study will be the first step toward distributed systems and will attempt to answer the following question: How to build fault tolerant distributed object storage and optimize it for small files?

The primary aims of the this research is following:

1. Design, implementation, and evaluation of an object storage system optimized for small files
2. Understanding the basics of the storage system
3. Understanding the basics of distributed systems

Data is needed to test the system's correctness. Primary data sources will be media files such as images, videos, and PDFs. For that, I have downloaded images and PDFs from the following resources, respectively: https://github.com/manjunath5496/Open-Access-Books and https://github.com/unsplash/datasets. Next, I will download a couple of videos from YouTube. If the video size is not small enough, I will divide it into smaller clips(30 - 60 seconds) for testing. This will not be a problem since the content of the media does not matter.

In image datasets, if some images are larger than normal (outliers), I will do some resizing or cropping to reduce the size. Additionally, for testing, I will create text files with the script.