

Simple Object Storage

Nijad Huseynov

May 2023

1 What are you going to do?

I propose to implement a simple object storage system. Object storage is simply a technology that stores unstructured data including but not limited to images, videos, PDFs, audio files, emails. This kind of storage systems are used in Big Data Analytics, Backup and archiving, Cloud applications, and so on [1, 2].

2 How is it done today? Current Limitations?

From the technical point of view, object storage systems are generally implemented in distributed fashion for scalability and fault tolerance. For example, Ceph - a well known distributed object storage, contains a metadata server which is responsible for storing various attributes of an objects and data server which is storing actual content of an object [3–5].

3 What is your idea to do something better?

By implementing this project, I want to gain practical experience in distributed object storage systems. My primary goal during the implementation will be to create a working example of object storage and hopefully learn about fundamental challenges and problems that arises such as guaranteeing consistency, scalability, performance, and others. After achieving my first goal, I will try to optimize the system for large number of small files.

4 Who will benefit from your work? Why?

Mostly it will be me. What the system does sounds simple. However, implementing a object storage is challenging task and requires to do lots of design decisions. In addition, it is an educational problem and can be a good project to practice fundamental computer science knowledge and software engineering principles.

5 What risks do you anticipate?

While implementing the object storage, I anticipate following problems.

- Technical implementation challenges. I don't have any prior experience in implementing any kind of storage system and I believe implementing such systems requires lots of researching, testing and deep understanding of distributed systems, operating systems, file systems, and others.
- Guaranteeing data consistency. In general, this is a hard problem and mechanism should be implemented and carefully tested.
- Failing to finish to project before the deadline. Depending on the complexities of the encountered problems during the implementation, finding a solution may take longer than I expected

6 Out of pocket costs? Complete within 11 weeks?

At this moment, I don't expect out of pocket costs.

7 Midterm results

By midterm, I plan to achieve following results

- Final System Design and Architecture. During the first weeks I will investigate existing object storage systems, research papers and conclude the final architecture.
- Implement some core functionalities such as creating an object, retrieving an object

8 Final Demonstration

At the end of the research class, I plan to demonstrate functional object storage and its features and how it is implemented. In addition I will evaluate the performance of the system and possibly compare it other existing open source solutions.

References

- [1] What is object storage? <https://cloud.google.com/learn/what-is-object-storage>.
- [2] What is object storage? <https://aws.amazon.com/what-is/object-storage>.
- [3] Sage A. Weil, Scott A. Brandt, Ethan L. Miller, Darrell D. E. Long, and Carlos Maltzahn. Ceph: A scalable, high-performance distributed file system. In *Proceedings of the 7th Symposium on Operating Systems Design and Implementation*, OSDI '06, page 307320, USA, 2006. USENIX Association.
- [4] Abutalib Aghayev, Sage Weil, Michael Kuchnik, Mark Nelson, Gregory R. Ganger, and George Amvrosiadis. The case for custom storage backends in distributed storage systems. *ACM Trans. Storage*, 16(2), may 2020.
- [5] Ceph architecture. <https://docs.ceph.com/en/latest/architecture/>.