2020 级

《大数据存储与管理》课程

实 验 报 告

姓　名　刘玺语

学　号　U202015299

班　号　计算机 2001 班

日　期　2023.05.29

目　　录

一、实验目的

1. 熟悉对象存储技术，代表性系统及其特性；
2. 实践对象存储系统，部署实验环境，进行初步测试；
3. 基于对象存储系统，分析性能问题，架设应用实践。

二、实验背景

随着信息技术普及，设备与用户数量急剧增长，当今互联网环境下每秒钟增长的数据量呈指数级增长，传统的文件系统和数据库无法应对日益增长的大数据存储管理需求(这其中也因为大数据背景下很多数据是非结构化的，SQL 传统关系数据库固定的模式无法应对这样的需求，因此 NoSQL 应运而生：文档数据库、k-v 数据库、图数据库等)。这也催生了管理非结构化数据的现代对象对出系统。根据设计好的协议，每个对象具有唯一 ID，对系统发起的请求是统一格式的，但得到的响应也许来自本地，也许来自云端，为存储系统的部署和管理带来了很强的灵活性。

非结构化数据必须以易于访问的方式来存储，兴起的对象存储技术综合了原网络存储技术的高速直接访问和数据共享等优势，同时也提供了具有高性能、高可靠性、跨平台以及安全的数据共享的存储体系结构。

三、实验环境

处理器 Intel(R) Core(TM) i7-10875H CPU @ 2.30GHz　2.30 GHz
机带 RAM 32.0 GB
系统类型　64 位操作系统, 基于 x64 的处理器

java version "17.0.4" 2022-07-19 LTS
Java(TM) SE Runtime Environment (build 17.0.4+11-LTS-179)
Java HotSpot(TM) 64-Bit Server VM (build 17.0.4+11-LTS-179, mixed mode, sharing)

minio version RELEASE.2023-04-20T17-56-55Z (commit-id=c61c4b71b26aad048b8f3abe0ee24547964fc49f)
Runtime: go1.20.3 windows/amd64
License: GNU AGPLv3 <https://www.gnu.org/licenses/agpl-3.0.html>
Copyright: 2015-2023 MinIO, Inc.

四、实验内容

安装 Java 环境。配置并架设 Minio 服务端，使用 s3bench 进行测试。

4.1 对象存储技术实践

从浏览器访问对应端口上传本地文件进行测试，结果如图 4.1 所示。

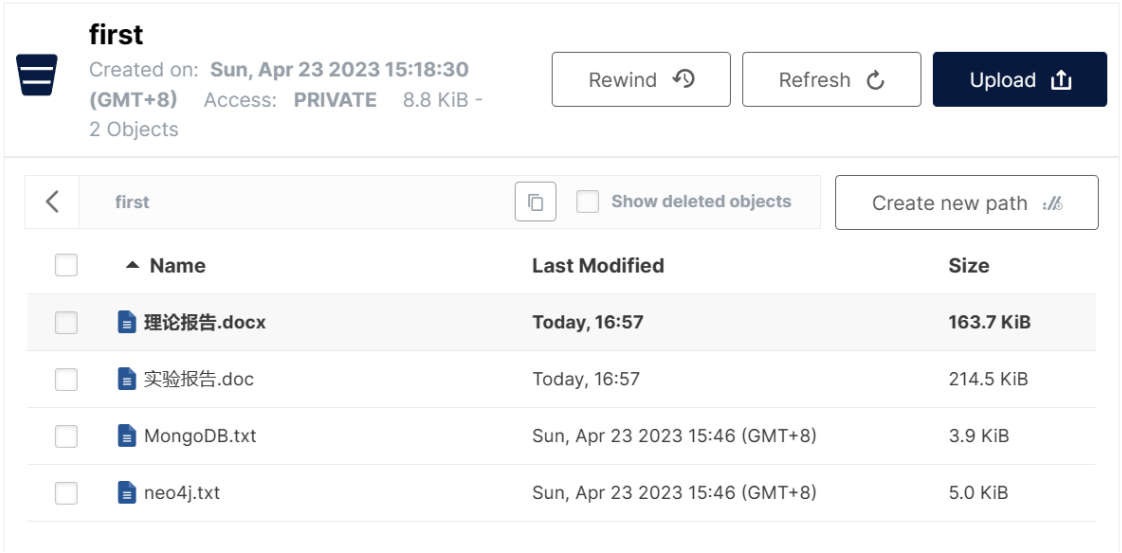

图 4.1 上传文件测试

4.2 对象存储性能分析

更改参数多次实验记录数据，excel 处理后制图如下(图 4.2，图 4.3).

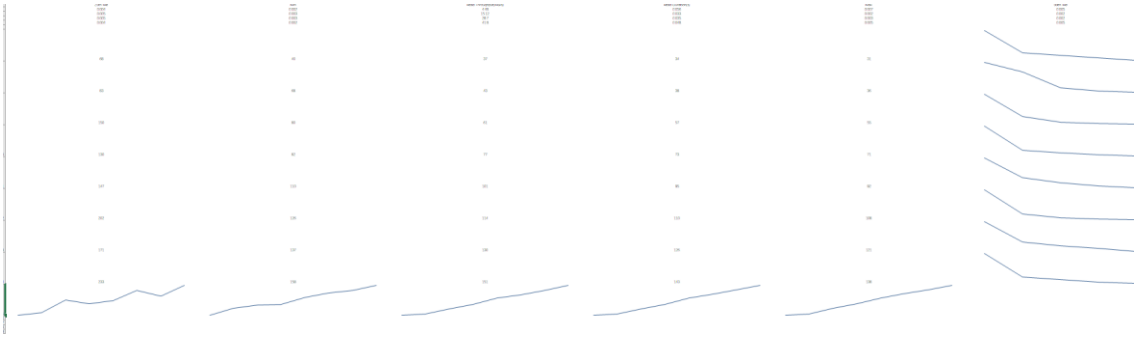

图 4.2 写入性能趋势图

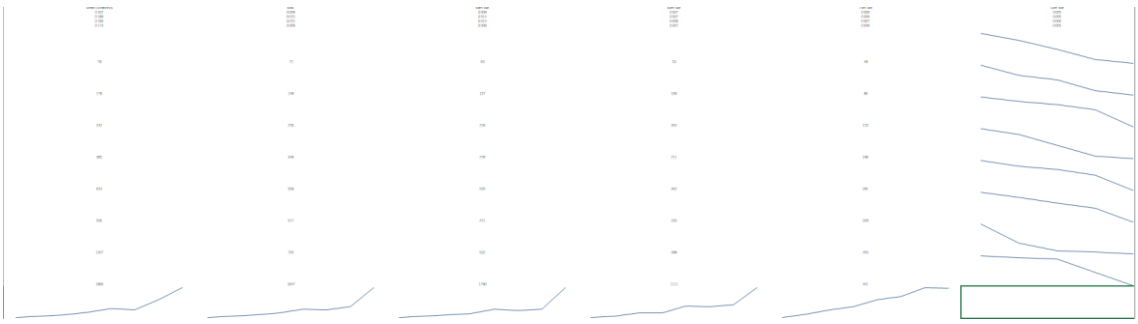

图 4.3 读取性能趋势图

五、实验过程

实验主要流程如图 5.1 和图 5.2 所示：

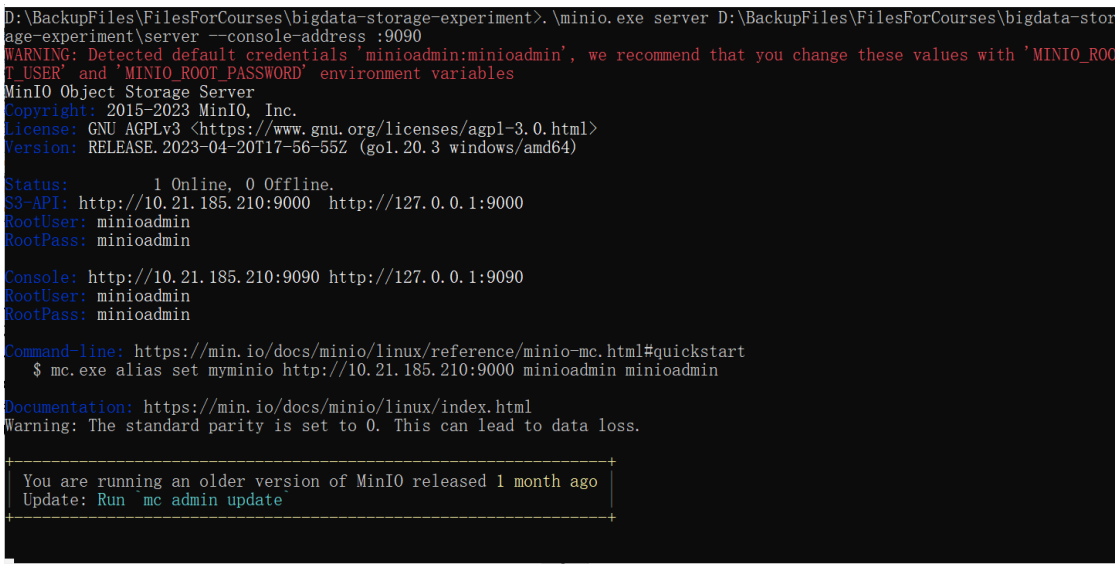

图 5.1 架设 minio 服务端

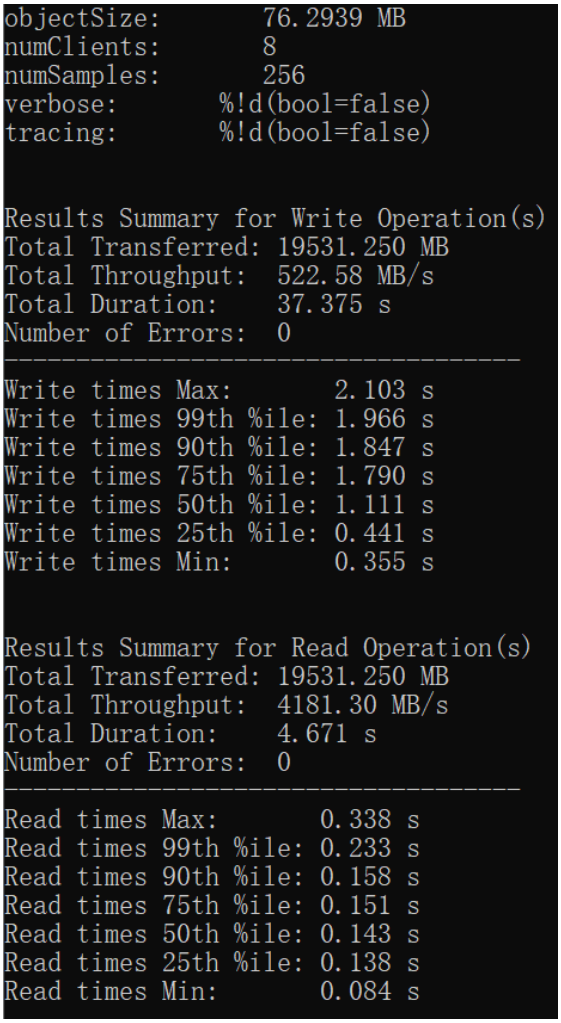

图 5.2 命令行调用 s3bench 测试

根据 4.2 对象存储性能分析中的图 4.2 和图 4.3 可见(竖直方向自上而下，对象大小为 10MB 到 80MB，水平方向从左至右 99%ile,90%ile,75%ile,50%ile,25%ile)。

随对象增大，存储用时增长约为线性增长到指数增长之间。

也可以从图中直观地看出尾延迟现象的存在(由于作图时是等距的，实际图像应更为平缓而后陡峭)。

从实验过程中发现，写成功率一直维持稳定，但写性能出现过不符合趋势的波动，但总体趋势未偏离理论预估值。

在 ObjectSize 较小时，read 容易出现读不成功。

六、实验总结

我在本次实验中收获颇丰。尤其是其中关于“计算机教育中缺失的一课”提到的相关知识与技能，在以往的编写代码过程中有些较少使用到的。如批处理脚本的语法和编写，同时也拓展学习了一些 git 的复杂功能。

在对于存储系统的使用和测试过程中，关于不同系统下的环境配置也吸取了很多经验。不过较为遗憾的是，WSL(Ubuntu22.04)环境下的 COSBench 并没有配置成功，多方搜寻资料也未能解决。

尽管如此还是在学习过程中拓展了许多知识面，以及关于尾延迟这一概念的理解和解决办法。

参考文献

[1] ZHENG Q, CHEN H, WANG Y 等. COSBench: A Benchmark Tool for Cloud Object Storage Services[C]//2012 IEEE Fifth International Conference on Cloud Computing. 2012: 998－999.

[2] ARNOLD J. OpenStack Swift[M]. O’Reilly Media, 2014.

[3] Dean J, Barroso L A. Association for Computing Machinery, 2013. The Tail at Scale[J]. Commun. ACM, 2013, 56(2): 74－80.

[4] Delimitrou C, Kozyrakis C. Association for Computing Machinery, 2018. Amdahl’s Law for Tail Latency[J]. Commun. ACM, 2018, 61(8): 65－72.