Building and Deploying a File Storage System using NextCloud

Introduction

Tasks and objectives:

- Implementing and deploying a cloud based file storage system
- Implementing user management and file operations
- Discussing security measures
- Discussing performances
- Discussing scalability

NextCloud key features

NextCloud comes with the following built-in features:

- Customizable database backend
- File operation for the users (upload, download, delete)
- User management and admin privileges
- Dockerized environment

Addressing Security

A number of measures can be taken to improve the security of the system:

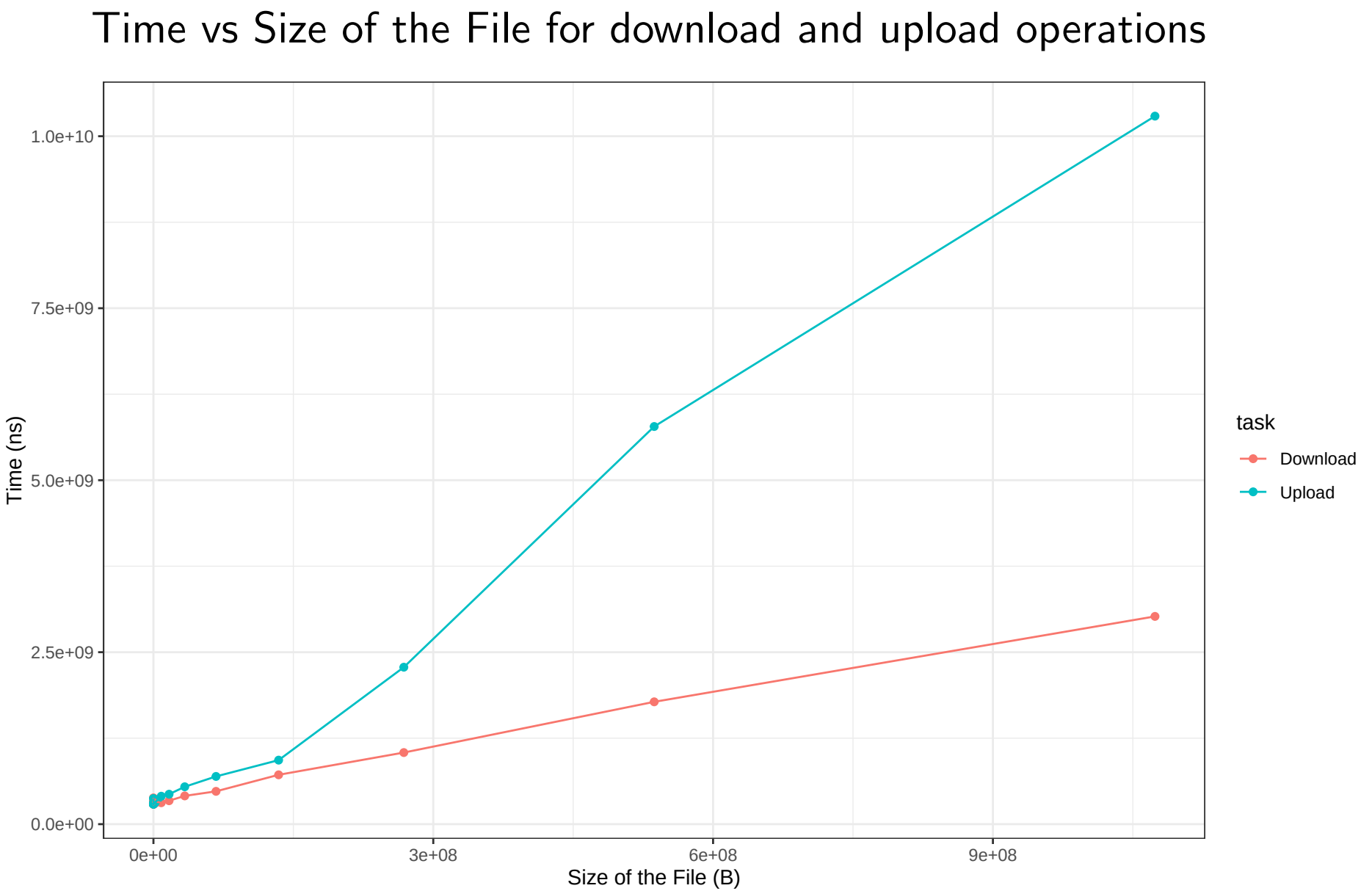
- Enable server-side data encryption (performance trade-off)
- Stronger password requirements (+ leaks check)
- Possibility to enable two factor authentication
- Possibility to enforce HTTPS protocol

Addressing Performance

Two different tools to address Performance:

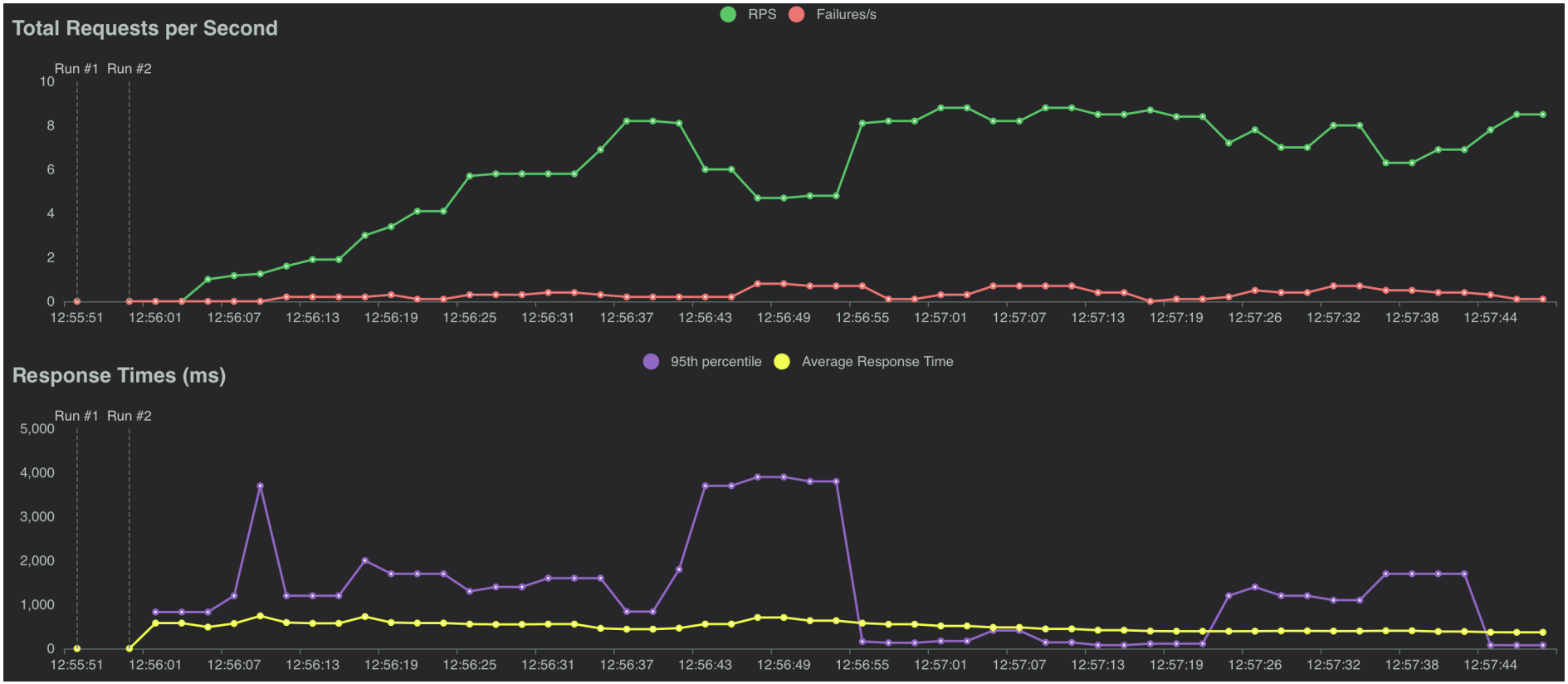
- **Curl**
  - command-line tool and library for transferring data
  - works with various network protocols, including HTTP and HTTPS
  - sends requests to a specified URL and receives response from the server
- **Locust**
  - open-source Python library and tool for performance testing and load testing of web applications
  - defines test scenarios, simulates user behavior, and generates load on a web server to perform stress-tests

Addressing Performance - Curl



Addressing Performance - Locust, 10 users

Addressing Performance - Locust, 30 users



Addressing Scalability

Two solutions for scaling the filesystem: cluster deployment and cloud services

- **Cluster Deployment:**
  - Deployment of NextCloud on a cluster of interconnected nodes.
  - Control over hardware resources and network configuration, customization and optimization in both performance and security.
  - High capex, Adhoc backup system, System Availability (mantainance downtime)
- **Cloud Services:**
  - Relies on cloud service providers to host and manage NextCloud instances in a cloud-based environment.
  - Scalability, flexibility, rapid deployment, automatic scaling, data redundancy, global accessibility.
  - Less control over infrastructure, possibly higher opex.