Cloud Computing

Basic and Advanced

Edoardo Zappia

September 2024

Basic part

Cloud-Based File Storage System

Objective: To design and implement a scalable, secure, and cost-efficient cloud-based file storage system

Tools Used:

- Nextcloud (file storage)
- MariaDB (database)
- Redis (cache)
- Docker, Docker-compose (containerization)
- Locust (load testing)

Cloud-Based File Storage System

Nextcloud Platform

Nextcloud is an open-source software that allows users to store files on their own private storage space

- Open-source
- User-friendly web interface
- Security and data privacy
- Compatibility with Docker
- Built-in features

Cloud-Based File Storage System

Deployment

Docker Deployment: docker-compose.yaml

- Nextcloud
- MariaDB database
- Redis caching
- Locust for load testing

Each service is configured with its respective docker image, environment variables, volumes for persistent storage and network settings

All containers share the same network

Cloud-Based File Storage System

User Authentication and File Operations

User Management:

- Role-based access control: regular users, admins
- User sign-up, log-in, and log-out
- Administrative user management

File operations and Storage:

- Private storage space
- Upload, download and delete files

Cloud-Based File Storage System

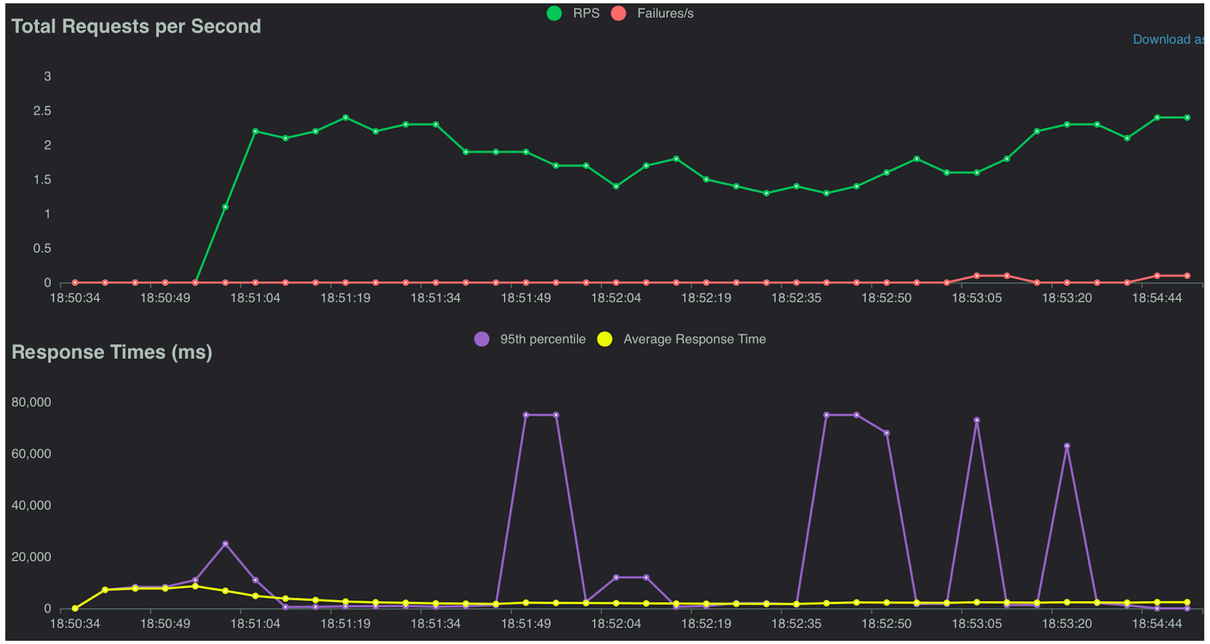
Security Measures

- **Two-Factor Authentication (2FA)**
  Different 2FA options (e.g., TOTP, SMS, hardware tokens)
- **Brute Force Protection**
- **File Access Control**
- **Password Policies**
- **Server-side encryption**

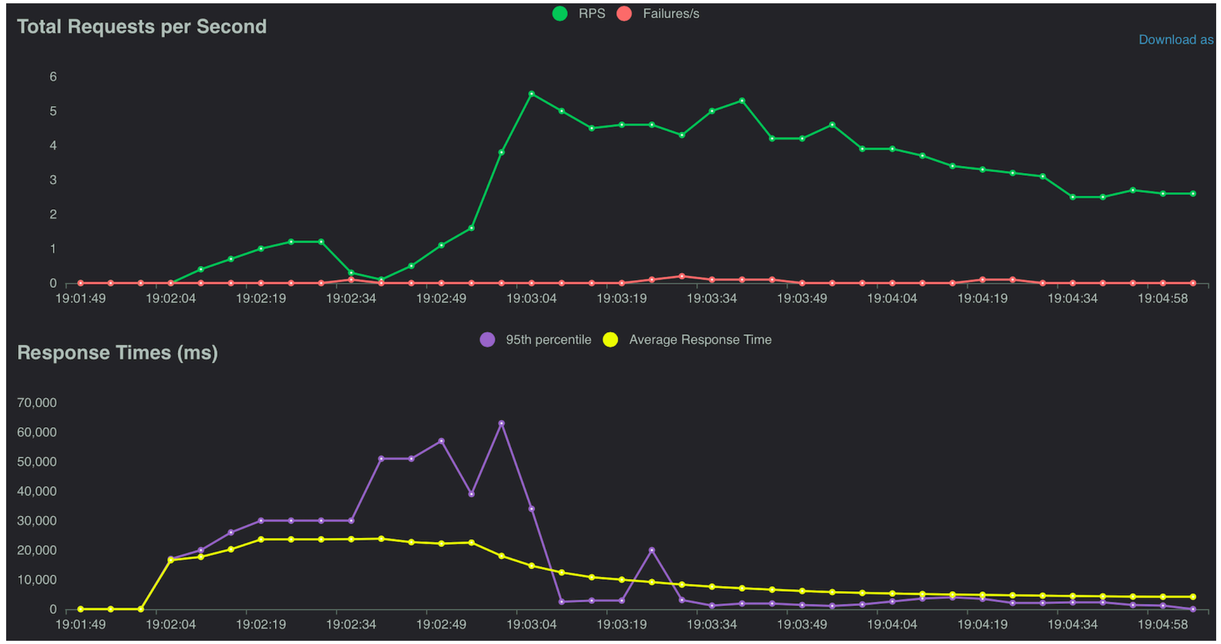
Cloud-Based File Storage System

Locust Test

10 users



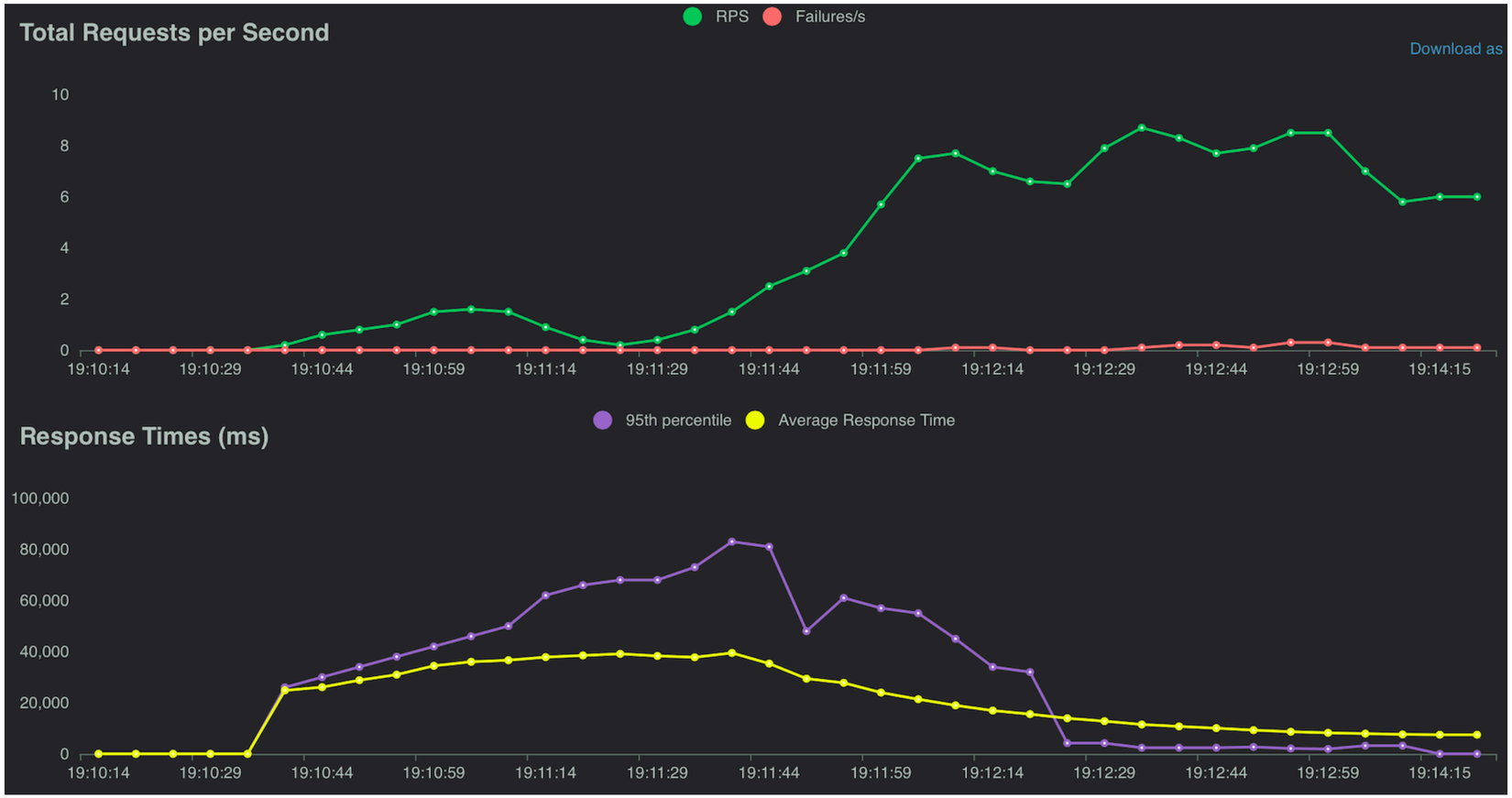
30 users



Cloud-Based File Storage System

Locust Test

50 users



Cloud-Based File Storage System

Scalability

Caching layers: to improve response time

Horizontal scaling: multiple Nextcloud instances

Object storage Service: to improve availability

Cloud-Based File Storage System

On-Premise Cluster

Advantages:

- hardware control
- data residency
- no third-party providers
- custom policies and security measures
- predictable costs, lower in the long run

Disadvantages:

- big initial investment
- maintenance and upgrades costs
- technical expertise

Cloud Providers

Amazon S3, Google Cloud Storage

	Amazon S3	Google Cloud Storage
Cost efficiency	Costs for storage per GB, requests, and data transfer with different tariff plans	Costs for storage per GB, requests, and data transfer with different tariff plans
Scalability	Designed to offer great scalability and handling of vast amounts of data and traffic	Google's private network provides fast data transfer rates
Security	Provides detailed access controls and integrates with AWS Identity and Access Management (IAM) for fine-grained security management	Google provides similar encryption capabilities and easy integration with Google Cloud's Identity and Access Management for permissions

Advanced part

Cloud-Based File Storage System in Kubernetes

Kubernets deployment: official Nextcloud Helm chart, values.yaml

- Single-node Kubernetes cluster
- Nextcloud pods with 3 containers, linked to a PVC:
  - Nextcloud application
  - sidecar container for cron jobs
  - nginx web server container
- External PostgreSQL database pod linked to a PVC
- Redis pod for caching
- MetalLB to expose the service
- Secrets

Cloud-Based File Storage System in Kubernetes

Back-End Storage Limitations

- **Single point of failure**
  Storage is not replicated across multiple nodes
- **Lack of dynamic provisioning**
  Manual PV creation required
- **Node affinity constraints**
  PV is tied to a specific node, the associated pods must run on the same node, limiting horizontal scalability
- **Limited scalability**
  Storage capacity is constrained by the physical storage available on the host

Cloud-Based File Storage System in Kubernetes

High Availability Considerations

- Adopt a distributed storage solution
- Enable dynamic provisioning
- Expand to a multi-node cluster
- Deploy multiple replicas
- Improve load balancing
- Horizontal pod autoscaler

Cloud-Based File Storage System in Kubernetes

Comparison with the Docker Solution

Advantages

- scalability
- high availability
- self-healing
- flexibility
- load balancing

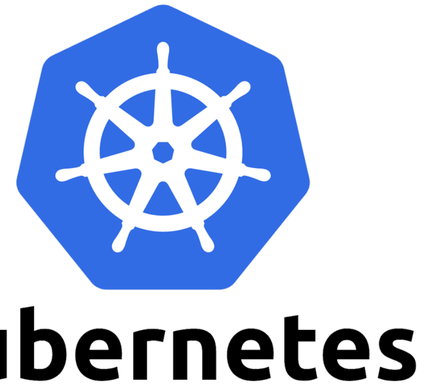
Disadvantages

- more complex to set up
- requires more resources

MPI service in Kubernetes

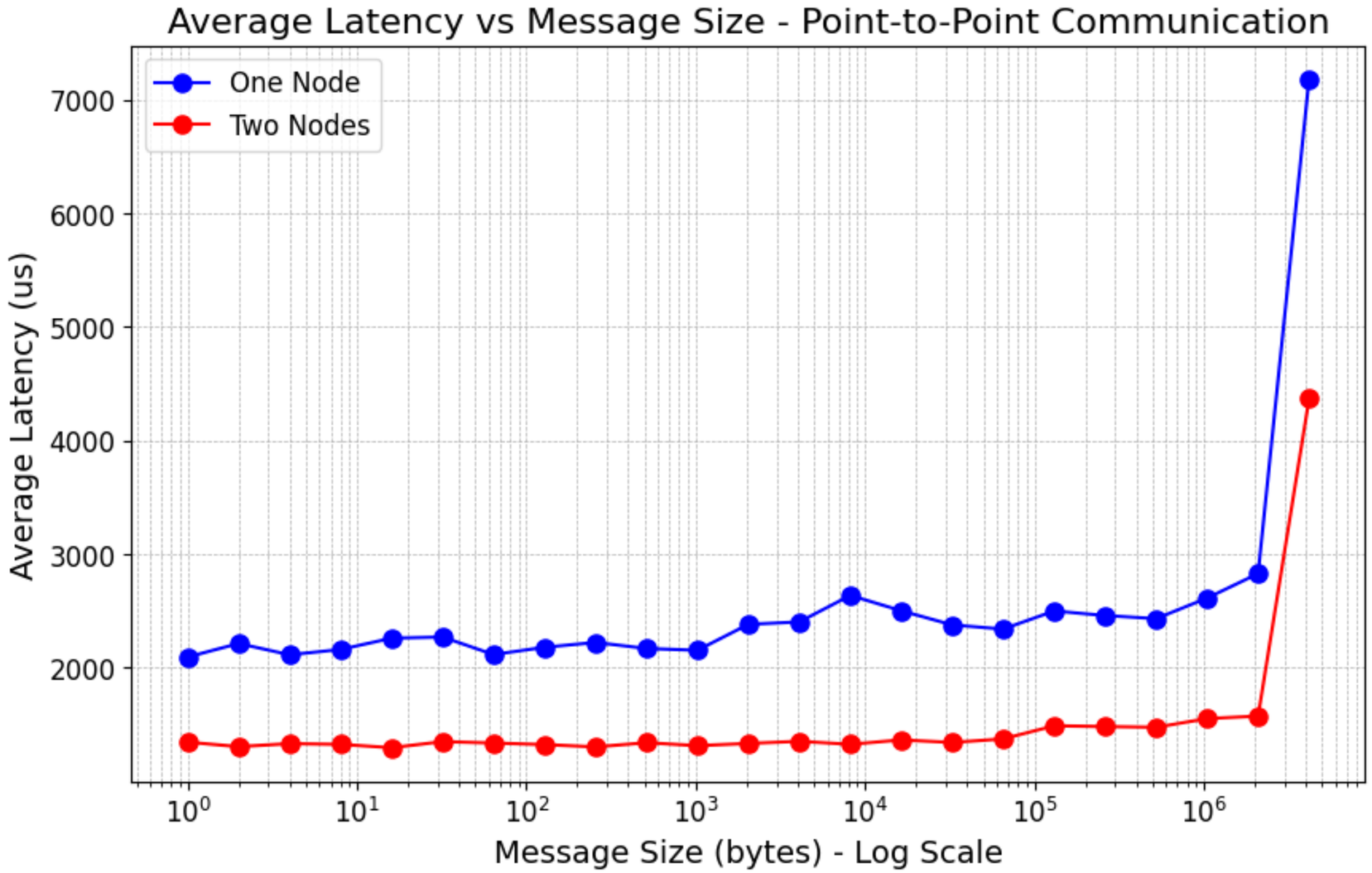
Structure

- Two-nodes Kubernetes cluster
- Flannel for communication between nodes
- MPI operator
- OSU benchmark container



MPI service in Kubernetes

OSU Benchmark: Point to Point Latency



MPI service in Kubernetes

OSU Benchmark: Scatter Collective Operation Latency