

**BỘ THÔNG TIN VÀ TRUYỀN THÔNG
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN
THÔNG**



BÁO CÁO THỰC TẬP CƠ SỞ

TÌM HIỂU VỀ DOCKER

Giảng viên hướng dẫn : Kim Ngọc Bách

Họ và tên : Ngô Vũ Minh Quý

Mã sinh viên : B22DCVT427

Lớp : E22CQCN02-B

MỤC LỤC

I. Giới thiệu chung về MinIO

II. Ứng dụng MinIO trong dự án Big Data

III. Hướng dẫn cài đặt miniIO

IV. Các thao tác và tích hợp cơ bản với MinIO

V. Đánh giá hiệu quả và kết luận

I. GIỚI THIỆU CHUNG VỀ MINIO



MinIO là một hệ thống lưu trữ đối tượng mã nguồn mở, nổi bật bởi khả năng tương thích hoàn toàn với Amazon S3 API. Điều này giúp các tổ chức dễ dàng chuyển đổi từ các dịch vụ lưu trữ đám mây sang sử dụng giải pháp tại chỗ hoặc tự quản lý trên nền tảng MinIO mà không cần viết lại ứng dụng.

MinIO được xây dựng theo triết lý đơn giản, hiệu quả và cloud-native. Điều này có nghĩa là nó có thể chạy tốt trong các môi trường container như Docker, Kubernetes và tích hợp dễ dàng vào các hệ thống CI/CD hiện đại.

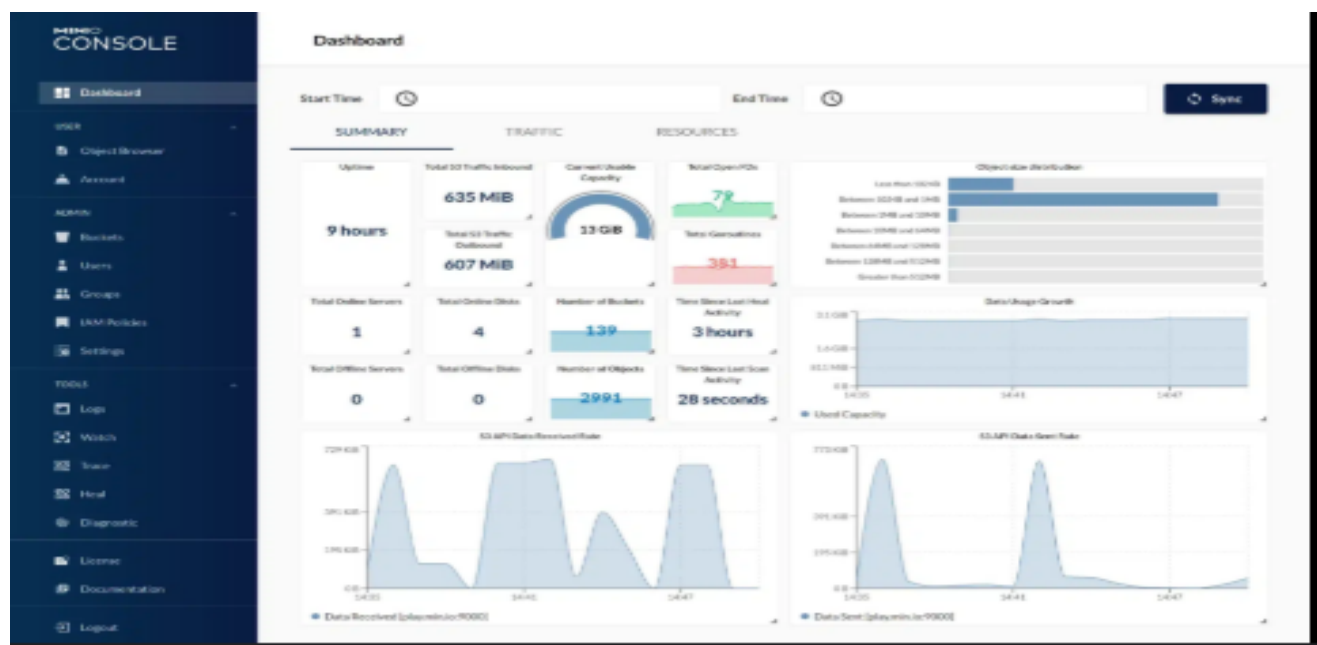
Các đặc điểm nổi bật của MinIO:

- **Mã nguồn mở 100%:** Dễ dàng kiểm tra, sửa đổi và triển khai.
- **Tương thích Amazon S3 API:** Cho phép sử dụng các SDK và công cụ đã quen thuộc.
- **Hiệu suất cao:** Có khả năng truyền tải dữ liệu lên đến hàng GB/s.
- **Triển khai đơn giản:** Hỗ trợ khởi chạy nhanh qua CLI, Docker hoặc Kubernetes.
- **Hỗ trợ bảo mật mạnh mẽ:** Bao gồm mã hóa dữ liệu và quản lý quyền truy cập chi tiết

MinIO được sử dụng phổ biến trong các hệ thống lưu trữ dữ liệu phi cấu trúc như hình ảnh, video, log, backup, big data, AI/ML model v.v.

II. Ứng dụng Docker trong phát triển phần mềm

MinIO là một giải pháp lưu trữ đối tượng được thiết kế để hoạt động hiệu quả trong các môi trường dữ liệu lớn (Big Data). Trong thời đại mà dữ liệu phi cấu trúc như log, hình ảnh, video, bản ghi cảm biến, và dữ liệu đầu vào từ hệ thống phân tán ngày càng tăng, MinIO đóng vai trò là một thành phần cốt lõi giúp đảm bảo việc lưu trữ, truy xuất và tích hợp dữ liệu một cách ổn định, bảo mật và linh hoạt.



1. Lưu trữ dữ liệu phi cấu trúc quy mô lớn

Các hệ thống Big Data thường phải xử lý một lượng dữ liệu khổng lồ được sinh ra liên tục từ nhiều nguồn. Đây thường là các loại dữ liệu không có cấu trúc rõ ràng như:

- File log hệ thống
- Ảnh, video, âm thanh từ thiết bị IoT
- Dữ liệu đầu vào từ cảm biến

- Tài liệu văn bản, JSON, XML..

MinIO hỗ trợ lưu trữ hiệu quả các loại dữ liệu này nhờ kiến trúc lưu trữ đối tượng (object storage), dễ dàng mở rộng từ vài terabyte đến hàng exabyte mà không ảnh hưởng đến hiệu suất.

2. Lưu trữ dữ liệu trung gian trong quy trình ETL

Trong quy trình xử lý dữ liệu kiểu ETL (Extract - Transform - Load), MinIO có thể được sử dụng để lưu trữ:

- Dữ liệu thô vừa được thu thập từ hệ thống nguồn
- Dữ liệu đã qua xử lý, phân loại, hoặc chuyển đổi định dạng
- Các tập dữ liệu tạm thời phục vụ phân tích hoặc học máy

Việc sử dụng MinIO giúp tách biệt rõ ràng giữa các bước trong pipeline, đảm bảo dữ liệu được lưu trữ bền vững và có thể truy xuất nhanh chóng trong các bước tiếp theo.

3. Tích hợp với các công cụ xử lý dữ liệu lớn

MinIO tương thích với **S3 API**, do đó có thể dễ dàng tích hợp với các nền tảng Big Data phổ biến như:

- **Apache Spark**: Đọc/ghi dữ liệu từ MinIO như thể từ S3, sử dụng định dạng Parquet, CSV, JSON...
- **Apache Hadoop**: Kết nối qua giao thức **s3a://** để lưu trữ hoặc nạp dữ liệu.
- **Presto, Hive, Trino**: Truy vấn trực tiếp dữ liệu lưu trong MinIO mà không cần thay đổi cấu trúc dữ liệu.
- **Flink, Kafka Connect**: Kết nối qua S3 Sink để ghi dữ liệu trực tiếp vào MinIO.

Điều này giúp các kỹ sư dữ liệu và nhà khoa học dữ liệu dễ dàng tích hợp MinIO vào các hệ thống phân tích lớn mà không cần viết lại nhiều đoạn mã hoặc cấu hình.

4. Lưu trữ kết quả và mô hình học máy

Trong lĩnh vực **trí tuệ nhân tạo và học máy (AI/ML)**, MinIO là nơi lưu trữ:

- Các tập dữ liệu huấn luyện lớn (training datasets)
- Kết quả sau huấn luyện: mô hình, trọng số, metadata...
- Kết quả suy luận từ hệ thống AI
- Log và metrics từ quá trình huấn luyện

III. HƯỚNG DẪN CÀI ĐẶT MINIO

+ Cài đặt MinIO trên hệ điều hành Linux / macOS

Bước 1: Tải xuống MinIO

```
bash                                                                    Sao chép  Chỉnh sửa
wget https://dl.min.io/server/minio/release/linux-amd64/minio
chmod +x minio
```

Bước 2: Khởi động MinIO Server

```
bash                                                                    Sao chép  Chỉnh sửa
./minio server /data
```

- **/data** là thư mục sẽ được sử dụng để lưu trữ dữ liệu. Có thể thay đổi thành thư mục tùy ý.
- Sau khi chạy lệnh, địa chỉ truy cập sẽ hiển thị trên terminal, thường là:
<http://localhost:9000>

Giao diện quản trị (MinIO Console): <http://localhost:9001> (nếu chỉ định **--console-address**)

Thông tin đăng nhập mặc định:

- Username: **minioadmin**
- Password: **minioadmin**

IV. Các thao tác cơ bản với Docker

Lưu ý: Trước khi sử dụng, cần cài đặt **mc** (MinIO Client) và cấu hình alias:

```
bash
# Tải và cài đặt mc
wget https://dl.min.io/client/mc/release/linux-amd64/mc
chmod +x mc

# Kết nối với MinIO local
./mc alias set myminio http://localhost:9000 minioadmin minioadmin
```

+Tải lên tệp (upload file)

```
bash
./mc cp example.txt myminio/mybucket
```

+ Tải xuống tệp (download file)

```
bash
./mc cp myminio/mybucket/example.txt ./downloaded-example.txt
```

+Liệt kê nội dung bucket

```
bash
./mc ls myminio/mybucket
```

+Xóa đối tượng (tệp)

```
bash Sao chép Chỉnh sửa
./mc rm myminio/mybucket/example.txt
```

+ Thiết lập quyền truy cập bucket

```
bash Sao chép Chỉnh sửa
./mc policy set download myminio/mybucket
```

V.Đánh giá và kết luận

Sau quá trình nghiên cứu và thực hành triển khai MinIO, em nhận thấy rằng MinIO không chỉ là một giải pháp lưu trữ đối tượng mã nguồn mở, mà còn là một công cụ quan trọng, đóng vai trò trung tâm trong các hệ thống dữ liệu hiện đại, đặc biệt là các hệ thống Big Data, AI/ML và microservices.

1. Ưu điểm nổi bật của MinIO

Một trong những điểm mạnh ấn tượng nhất của MinIO là **hiệu suất xử lý dữ liệu cao**. Với khả năng đọc và ghi dữ liệu lên tới hàng GB/s, MinIO hoàn toàn đáp ứng tốt yêu cầu của các ứng dụng lớn, đặc biệt là những ứng dụng cần xử lý dữ liệu phi cấu trúc liên tục như hệ thống log, cảm biến IoT, hoặc dữ liệu video.

Khả năng mở rộng linh hoạt cũng là điểm cộng lớn của MinIO. Nó có thể chạy đơn giản trên một máy chủ, nhưng cũng có thể mở rộng thành hệ thống phân tán để xử lý dữ liệu lên đến hàng exabyte. Điều này giúp người dùng dễ dàng bắt đầu từ quy mô nhỏ và phát triển dần theo nhu cầu.

2. Ứng dụng thực tế trong hệ thống Big Data

Qua việc ứng dụng MinIO trong các dự án mô phỏng như hệ thống thu thập và phân tích log thời gian thực, em nhận thấy vai trò của MinIO rất rõ ràng

trong nhiều khía cạnh:

- **Lưu trữ log thô** với độ tin cậy cao, phục vụ mục đích phân tích, kiểm tra, và truy vết hệ thống.
- **Tích hợp với Spark hoặc Flink** để xử lý dữ liệu trực tiếp từ MinIO thông qua S3 API, đơn giản hóa pipeline xử lý.
- **Lưu trữ kết quả phân tích và mô hình học máy**, đảm bảo tính linh hoạt và khả năng tái sử dụng.
- **Kết nối với dashboard trực quan (như Grafana)** để theo dõi hệ thống theo thời gian thực.

Việc triển khai MinIO giúp tách biệt các bước trong chuỗi xử lý dữ liệu, đồng thời giúp hệ thống dễ mở rộng, dễ bảo trì và dễ quản lý hơn.

3. Đánh giá tổng quan

MinIO thực sự là một giải pháp **đơn giản nhưng mạnh mẽ**, dễ tiếp cận cho cả sinh viên, nhà phát triển phần mềm và kỹ sư dữ liệu. Chỉ với vài dòng lệnh hoặc cấu hình Docker Compose, người dùng có thể nhanh chóng thiết lập hệ thống lưu trữ hiệu quả, bảo mật và ổn định.

Chi phí thấp là một lợi thế đáng kể khi so sánh với các dịch vụ đám mây thương mại. Điều này đặc biệt phù hợp với các tổ chức giáo dục, startup, hoặc dự án nghiên cứu cần kiểm soát chi phí chặt chẽ.