**TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI**



**VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**

─────── \* ───────



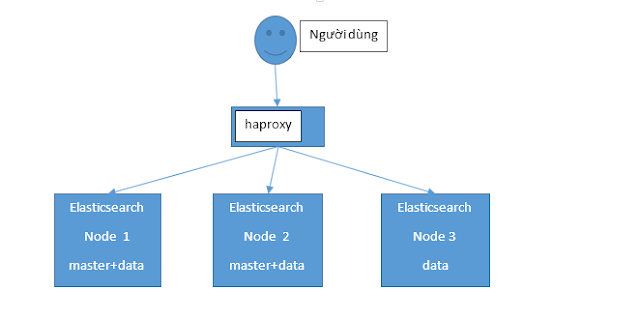
**BÀI TẬP TRÊN LỚP**

**MÔN HỌC: LƯU TRỮ VÀ XỬ LÝ DỮ LIỆU LỚN**

**LAB 2: ELASTICSEARCH**

NHÓM: GIỮA CHÚNG TA

**Hà Nội, 11-2020**



*Sơ đồ cụm cluster Elasticsearch*

Chúng ta sẽ cài đặt elasticsearch trên 3 máy chủ. Node 1 và node 2 làm master node, còn node 3 là data node.

**I) Cài đặt ElasticSearch trên 3 máy:**

**- Bước 1:** Luôn cập nhật hệ thống trước khi cài đặt

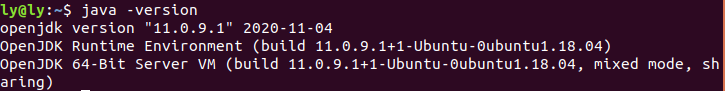
sudo apt-get update && apt-get upgrade

**- Bước 2:** Cài đặt JDK

sudo apt-get install default-jdk

Sau khi cài JDK xong, chúng ta có thể kiểm tra xem đã cài đặt thành công hay chưa bằng cách kiểm tra phiên bản JDK vừa cài bằng lệnh:

java -version



Nếu hiển thị như trong ảnh tức là đã cài đặt thành công.

**- Bước 3: Cài đặt Elasticsearch**

Chạy lần lượt các lệnh sau:

sudo apt-get install apt-transport-https

wget -qO - https://artifacts.elastic.co/GPG-KEY-elasticsearch | sudo apt-key add -

sudo add-apt-repository "deb https://artifacts.elastic.co/packages/5.x/apt stable main"

sudo apt-get update

sudo apt-get install elasticsearch

**- Bước 4: Config**

Mở file elasticsearch.yml bằng lệnh:

sudo nano /etc/elasticsearch/elasticsearch.yml

Chỉnh sửa như sau:

network.host: localhost

http.port: 9200

**- Bước 5: Khởi động Elasticsearch**

Chạy lần lượt các lệnh:

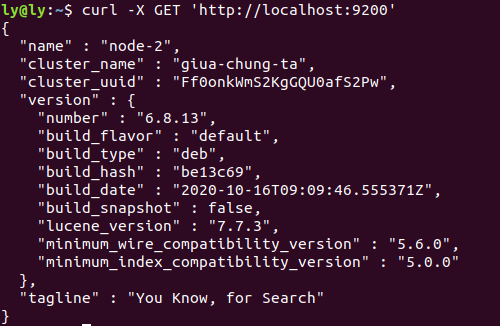
sudo /bin/systemctl enable elasticsearch.service

sudo systemctl start elasticsearch.service

Sau đó, ta có thể kiểm tra xem Elasticsearch đã được cài thành công trên máy hay chưa, nếu có hiển thị như ảnh dưới đây tức ta đã cài đặt thành công:

Sử dụng lệnh:

curl -X GET ‘localhost:9200’



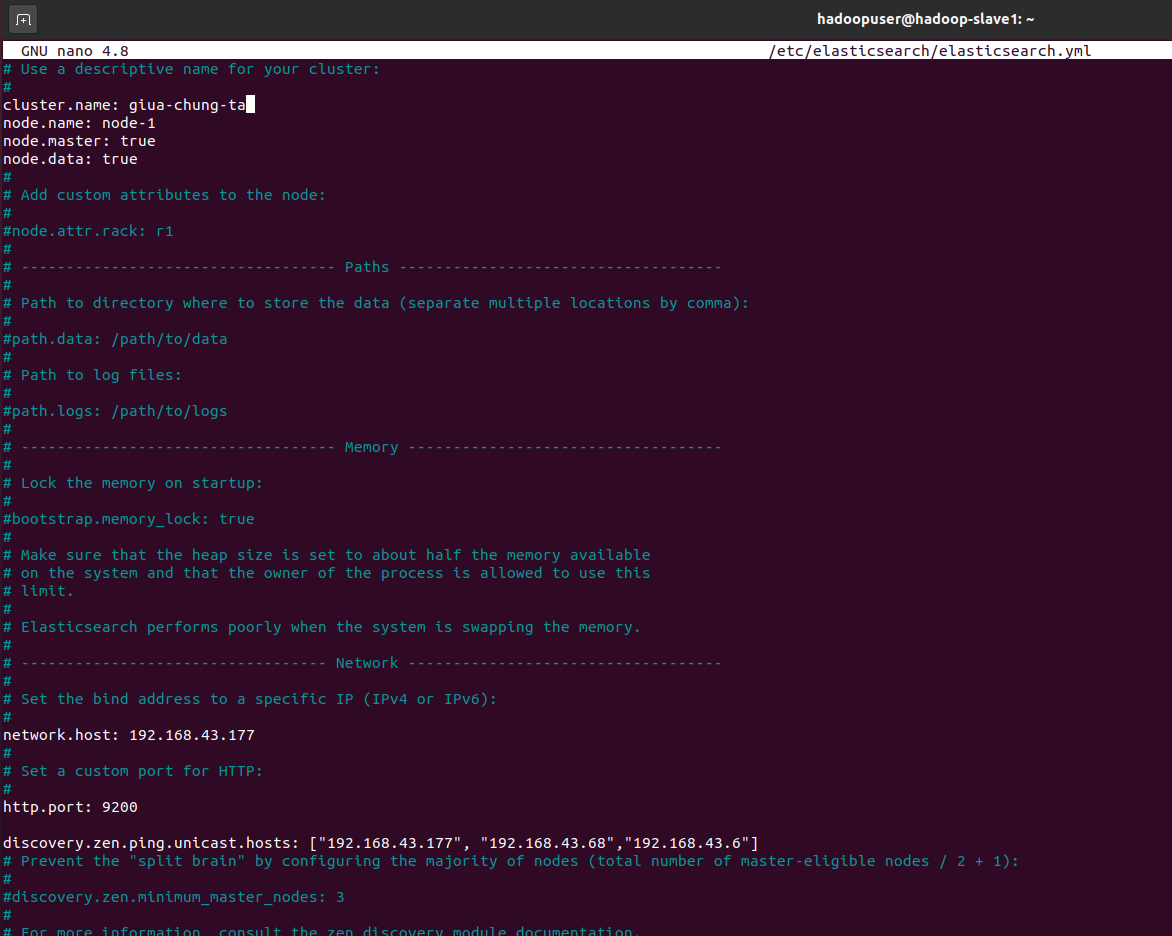
Làm lần lượt tất cả các bước trên cho cả 3 máy.

**II) Tạo cụm Elasticsearch:**

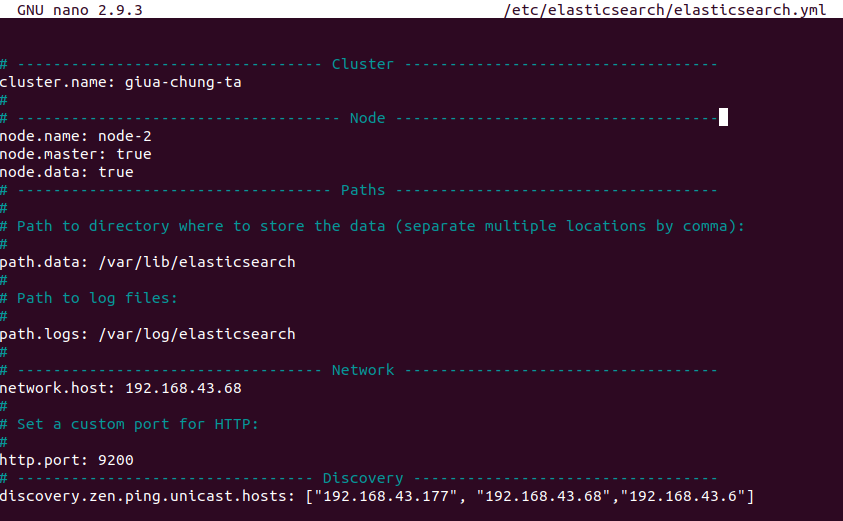
Chúng ta có 3 node: Node 1 và node 2 làm master node, node 3 làm data node. Đối với mỗi node, chúng ta đều cần mở tệp cấu hình Elasticsearch bằng câu lệnh sau:

sudo nano /etc/elasticsearch/elasticsearch.yml

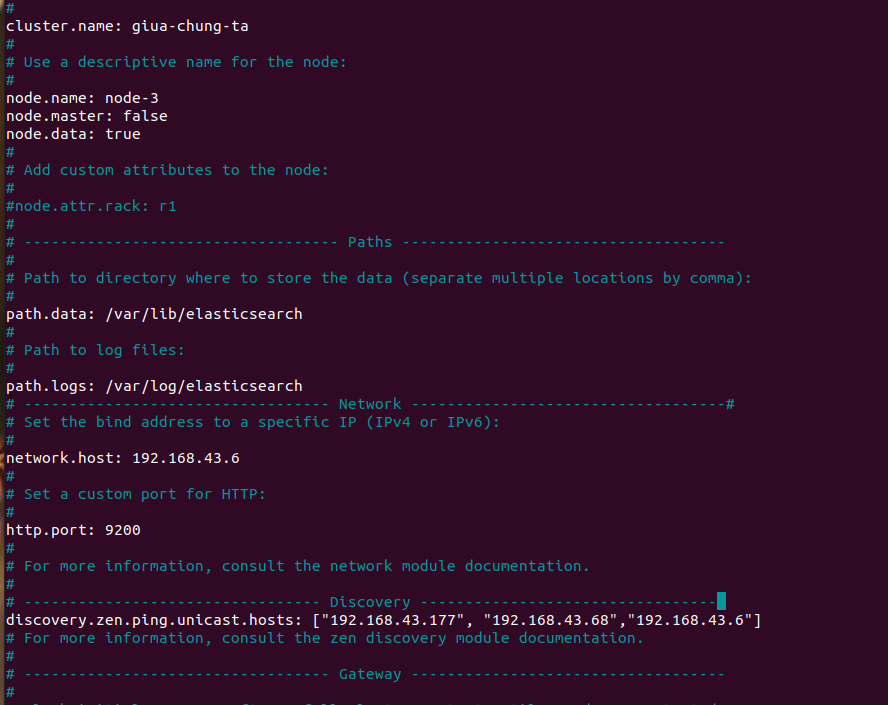
Đối với node 1, chúng ta cấu hình như sau:



Node 2 cấu hình như sau:



Node 3:



cluster-name: tên của cụm cluster, cấu hình cluster-name ở cả 3 máy phải giống nhau

network-host: điền địa chỉ IP của máy.

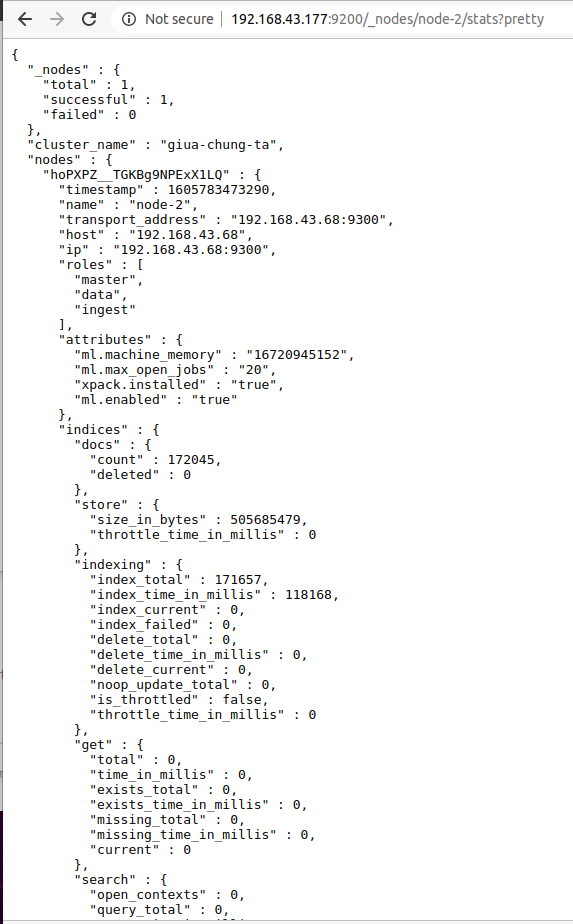
Sau đó, tại mỗi node, chúng ta chạy lệnh sau:

sudo service elasticsearch start

Nếu tất cả mọi thứ đã được cấu hình chính xác, cụm Elaticsearch của bạn sẽ hoạt động. Để xác minh mọi thứ đang hoạt động như mong đợi, hãy truy vấn Elaticsearch từ bất kỳ node nào bằng lệnh dưới đây:

curl -XGET 'http://[địa chỉ IP của bạn]:9200/\_cluster/state?pretty'

Kết quả thu được sẽ là data JSON giống như hình dưới:



**III) Upload 1GB data lên Elasticsearch:**

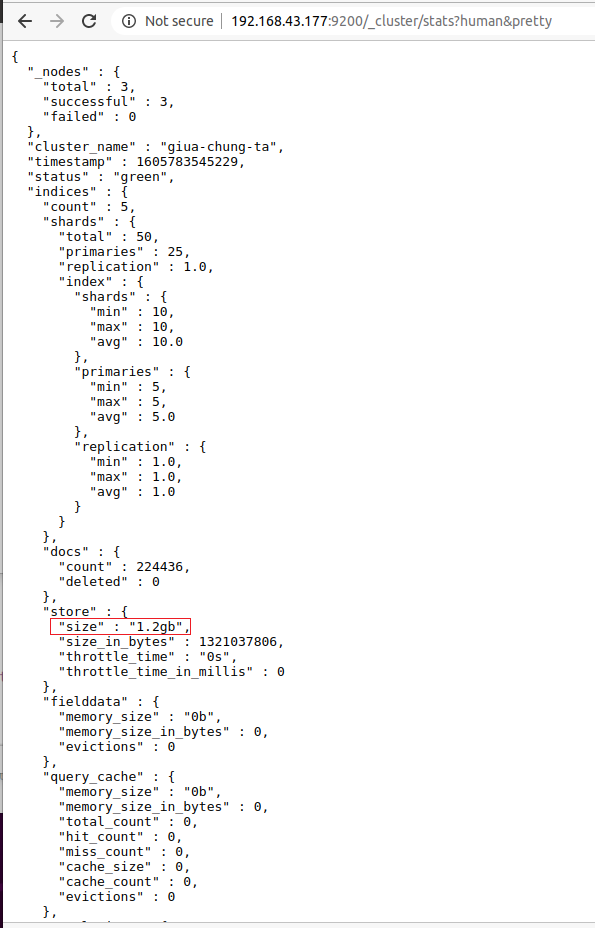
Chúng ta chuẩn bị 1GB file dữ liệu JSON.

Sau đó, cd đến thư mục lưu dữ liệu cần đẩy lên Elasticsearch.

Sau đó sử dụng câu lệnh sau để đưa dữ liệu vào Elasticsearch:

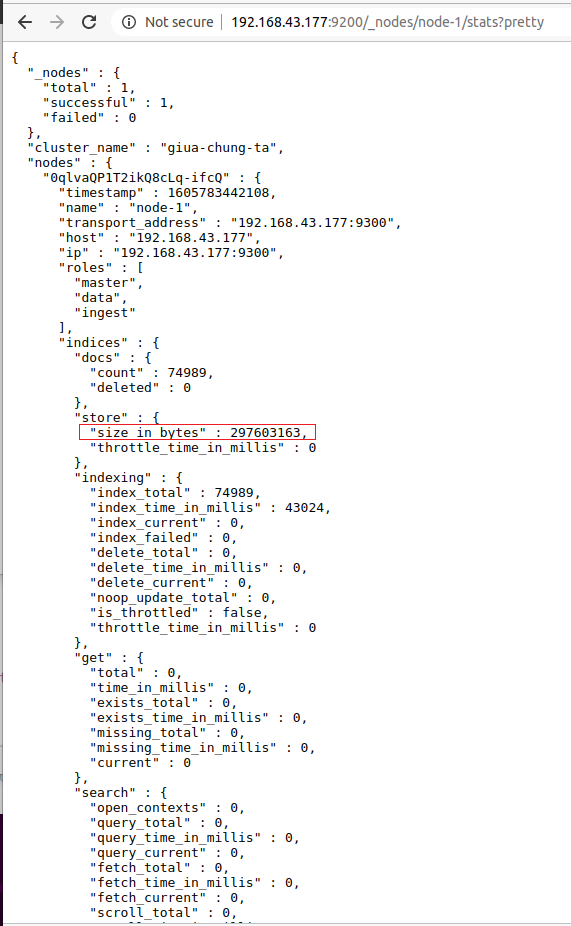
curl -H 'Content-Type: application/x-ndjson' -XPOST '[địa chỉ IP của bạn]:9200/[thư mục lưu trữ]/doc/\_bulk?pretty' --data-binary @[tên file]

Minh chứng đã upload 1GB dữ liệu:

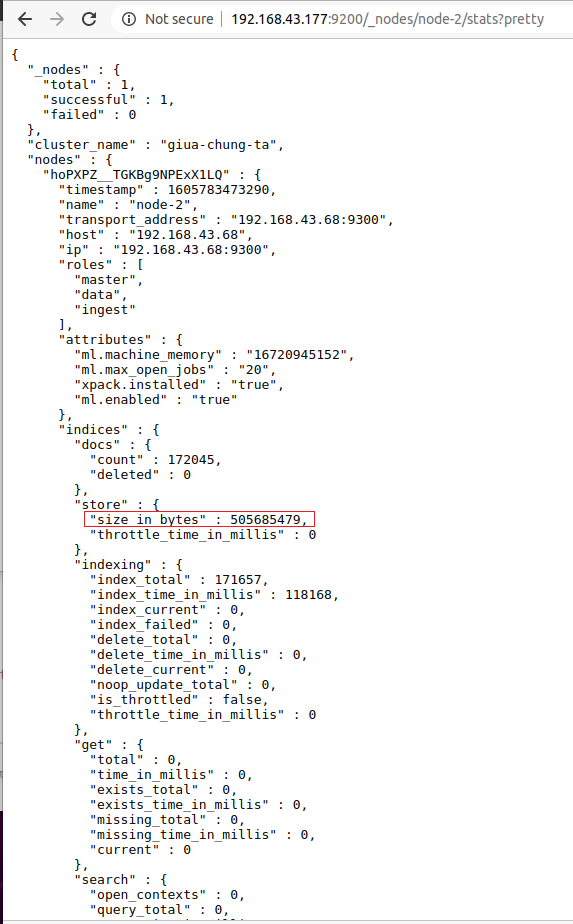


Minh chứng dữ liệu được lưu phân tán:

Trên node 1:



Trên node 2:



Trên node 3:

