

Hướng dẫn cài đặt và cấu hình Elasticsearch trên CentOS 7

Elasticsearch là một nền tảng cho việc tìm kiếm và phân tích dữ liệu phân tán trong thời gian thực. Để sử dụng, tính năng mạnh mẽ, và khả năng mở rộng.

Bạn có thể sử dụng các phương thức (GET, POST, PUT, DELETE,...) kết hợp với HTTP URI (/collection/entry) để thao tác dữ liệu. Các cách tiếp cận trực quan RESTful là cả nhà phát triển và người dùng thân thiện, đó là một trong những lý do Elasticsearch phổ biến.

Elasticsearch là một phần mềm mã nguồn mở và miễn phí với một công ty vững chắc đằng sau. Sự kết hợp này thích hợp để sử dụng trong bất cứ nơi nào từ thử nghiệm cá nhân để hội nhập của công ty.

Bài viết này sẽ giới thiệu đến Elasticsearch, cách cài đặt, cấu hình, và bắt đầu sử dụng.

Bước 1 — Cài đặt Java

- Java Runtime Environment (JRE)
- Cài đặt OpenJDK mới nhất với các lệnh:

```
sudo yum install java-1.8.0-openjdk.x86_64
```

```
java -version
```

Bước 2 — Downloa và cài Elasticsearch

Elasticsearch có thể được tải về trực tiếp từ [elastic.co](https://download.elastic.co) trong gói zip, tar.gz, deb, rpm.

```
Wget https://download.elastic.co/elasticsearch/elasticsearch/elasticsearch-1.7.3.noarch.rpm
```

Sau đó cài đặt nó trong CentOS cách thông thường bằng lệnh **rpm** như sau:

```
sudo rpm -ivh elasticsearch-1.7.3.noarch.rpm
```

Kết quả Elasticsearch được cài đặt trong `/usr/share/elasticsearch/` với các tập tin cấu hình của nó được đặt trong `/etc/elasticsearch` và có init script thêm vào trong `/etc/init.d/elasticsearch`.

Để đảm bảo Elasticsearch bắt đầu và dừng tự động với những Droplet, thêm init script để runlevel mặc định bằng lệnh:

```
sudo systemctl enable elasticsearch.service
```

Bước 3 — Cấu hình Elastic

Các tập tin cấu hình Elasticsearch có trong `/etc/elasticsearch`. Có 2 file:

- `elasticsearch.yml` — Cấu hình máy chủ Elasticsearch.
- `logging.yml` — Cung cấp cấu hình cho logging. Có thể để các tùy chọn logging mặc định. Bạn có thể tìm thấy các bản ghi kết quả trong `/var/log/elasticsearch` by default.

Các biến đầu tiên để tùy chỉnh trên một máy chủ Elasticsearch là `node.name` và `cluster.name` trong `elasticsearch.yml`. Như tên gợi ý, `node.name` chỉ định tên của máy chủ (node) và các cụm mà sau này có liên quan.

Nếu bạn không tùy chỉnh các biến, một `node.name` sẽ được tự động đặt trong mối liên quan với hostname Droplet. Các `cluster.name` sẽ được tự động thiết lập để tên của cluster mặc định.

Giá trị `cluster.name` được sử dụng bởi tính năng tự động phát hiện Elasticsearch để tự động phát hiện và kết Elasticsearch các nút cho một cluster. Vì vậy, nếu bạn không thay đổi giá trị mặc định, bạn có thể có các nút không mong muốn, được tìm thấy trên cùng một mạng, trong cluster của bạn.

Chỉnh sửa các tập tin cấu hình `elasticsearch.yml`:

```
sudo nano /etc/elasticsearch/elasticsearch.yml
```

Xóa dấu # đầu dòng `node.name` và `cluster.name`, và sau đó thay đổi giá trị của họ.

Cấu hình đầu tiên của bạn thay đổi trong

`/etc/elasticsearch/elasticsearch.yml` tập tin sẽ giống như thế này:

```
/etc/elasticsearch/elasticsearch.yml
...
node.name: "My First Node"
cluster.name: mycluster1
...
```

The setting which determines the role of the server is called `node.master`. If you have only one Elasticsearch node, you should leave this option commented out so that it keeps its default value of `true`— i.e. the sole node should be also a master.

Alternatively, if you wish to configure the node as a slave, remove the # character at the beginning of the `node.master` line, and change the value to `false`:

```
/etc/elasticsearch/elasticsearch.yml
...
node.master: false
...
```

Một tùy chọn cấu hình quan trọng là `node.data`, trong đó xác định liệu một nút sẽ lưu trữ dữ liệu hay không. Trong hầu hết các trường hợp tùy chọn này nên để cho giá trị mặc định của nó (`true`), nhưng có hai trường hợp mà bạn có thể muốn không để lưu trữ dữ liệu trên một nút. Một là khi các nút là một chuyên dụng "master", như chúng ta đã đề cập. Khác là khi một nút chỉ được sử dụng cho việc lấy dữ liệu từ các nút và tổng hợp kết quả. Trong trường hợp sau, nút sẽ đóng vai trò như một "cân bằng tải tìm kiếm".

Nếu bạn chỉ có một nút Elasticsearch, bạn nên để thiết lập commented cho nó giữ giá trị thực sự mặc định. Nếu không, để vô hiệu hóa việc lưu trữ dữ liệu tại locall, bỏ ghi chú dòng sau và thay đổi giá trị `false`:

```
/etc/elasticsearch/elasticsearch.yml
...
node.data: false
...
```

Hai tùy chọn quan trọng khác là

`index.number_of_shards` và `index.number_of_replicas`.

Giả sử rằng bạn vẫn đang khám phá và thử nghiệm Elasticsearch vào một nút duy nhất, tốt hơn để bắt đầu với chỉ một và không có bản sao. Như vậy, giá trị được thiết lập như sau:

```

/etc/elasticsearch/elasticsearch.yml
...
index.number_of_shards: 1
index.number_of_replicas: 0
...
```

Một thiết lập cuối cùng mà bạn có thể quan tâm trong việc thay đổi là `path.data`, mà quyết định đường dẫn, nơi dữ liệu được lưu trữ. Đường dẫn mặc định là `/var/lib/elasticsearch`.

Nên sử dụng một phân vùng riêng và điểm gắn kết để lưu trữ dữ liệu Elasticsearch. Trong trường hợp tốt nhất, phân vùng dành riêng này sẽ là một phương tiện lưu trữ riêng biệt mà sẽ cung cấp hiệu suất tốt hơn và dữ liệu tách biệt. Bạn có thể chỉ định đường dẫn khác nhau `path.data` bỏ `#` dòng `path.data` và thay đổi giá trị của nó:

```

/etc/elasticsearch/elasticsearch.yml
...
path.data: /media/different_media
...
```

Bắt đầu chạy Elasticsearch cho lần đầu tiên với lệnh:

```
sudo service elasticsearch start
```

Bước 4 — Bảo mật Elastic

Elasticsearch không có built-in bảo mật và có thể được kiểm soát bởi bất cứ ai có thể truy cập vào các API, HTTP. Vì vậy, các tinh chỉnh bảo mật đầu tiên là để ngăn chặn truy cập. Để loại bỏ quyền truy cập chỉnh sửa các tập tin `elasticsearch.yml`:

```
sudo nano /etc/elasticsearch/elasticsearch.yml
```

Tìm dòng có chứa `network.bind_host`, bỏ ghi chú nó bằng cách loại bỏ các ký tự `#` ở đầu dòng, và thay đổi giá trị `localhost` tương tự như sau:

```
/etc/elasticsearch/elasticsearch.yml
...
network.bind_host: localhost
...
```

Để vô hiệu hóa các biểu thức tùy chỉnh, thêm dòng sau đây là ở phần cuối của file `/etc/elasticsearch/elasticsearch.yml`:

```
/etc/elasticsearch/elasticsearch.yml
...
script.disable_dynamic: true
...
```

Khởi động lại Elasticsearch với lệnh:

```
sudo service elasticsearch restart
```

Bước 5 — Testing

Elasticsearch nên được chạy trên cổng 9200.

```
curl -X GET 'http://localhost:9200'
```

Bạn sẽ thấy các phản hồi sau:

Output of curl

```
{
  "status" : 200,
  "name" : "CentOS Node",
  "cluster_name" : "mysqlcluster",
  "version" : {
    "number" : "1.7.3",
    "build_hash" : "05d4530971ef0ea46d0f4fa6ee64dbc8df659682",
    "build_timestamp" : "2015-10-15T09:14:17Z",
    "build_snapshot" : false,
    "lucene_version" : "4.10.4"
  },
  "tagline" : "You Know, for Search"
}
```

Nếu một phản hồi tương tự như trên, Elasticsearch đang làm việc đúng cách. Nếu không, hãy chắc chắn rằng bạn đã làm theo đúng các hướng dẫn cài đặt.

Bước 6 — Sử dụng Elasticsearch

Để bắt đầu sử dụng Elasticsearch, hãy thêm một vài dữ liệu đầu tiên. Như đã đề cập, Elasticsearch sử dụng một API RESTful, đáp ứng các lệnh CRUD thông thường: Create, Read, Update và Delete. Chúng ta sẽ sử dụng curl một lần nữa.

```
curl -X POST 'http://localhost:9200/tutorial/helloworld/1' -d '{
  "message": "Hello World!" }'
```

Bạn sẽ thấy các phản hồi sau:

Output

```
{"_index":"tutorial","_type":"helloworld","_id":"1","_version":1,"created":true}
```

Với curl, ta đã gửi một yêu cầu HTTP POST cho máy chủ Elasticsearch. URI được yêu cầu là /tutorial/helloworld/1. Điều quan trọng là phải hiểu các thông số ở đây:

- tutorial là chỉ số của dữ liệu trong Elasticsearch.
- helloworld là kiểu.
- 1 là id của entry theo các chỉ số trên và kiểu.

Bạn có thể truy nhập lần đầu tiên với một yêu cầu HTTP GET như thế này:

```
curl -X GET 'http://localhost:9200/tutorial/helloworld/1'
```

Kết quả như sau:

Output

```
{"_index":"tutorial","_type":"helloworld","_id":"1","_version":1,"found":true,"_source":{"message": "Hello World!"}}
```

Để chỉnh sửa một entry hiện tại bạn có thể sử dụng một yêu cầu HTTP PUT như thế này:

```
curl -X PUT 'localhost:9200/tutorial/helloworld/1?pretty' -d '
```

```
{  
  
  "message": "Hello People!"  
  
}
```

Elasticsearch should acknowledge successful modification like this:

Output

```
{  
  "_index" : "tutorial",  
  "_type" : "helloworld",  
  "_id" : "1",  
  "_version" : 2,  
  "created" : false  
}
```

Bạn cũng có thể "tô điểm" kết quả của bạn khi truy xuất dữ liệu và nhận được kết quả tốt hơn nhiều như sau:

```
curl -X GET 'http://localhost:9200/tutorial/helloworld/1?pretty'
```

Kết quả:

Output

```
{  
  "_index" : "tutorial",  
  "_type" : "helloworld",  
  "_id" : "1",  
  "_version" : 2,  
  "found" : true,  
  "_source":{ "message": "Hello World!" }  
}
```

Trên đây là cách để cài đặt, cấu hình và bắt đầu sử dụng Elasticsearch.