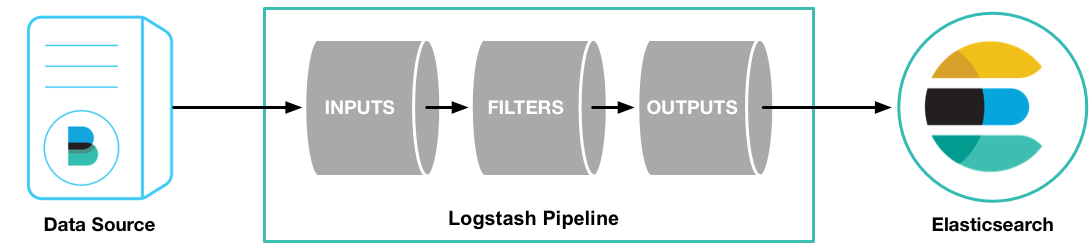
**LOGSTASH**

**1. Giới thiệu**

- Logstash là một công cụ thu thập dữ liệu mã nguồn mở với khả năng tổng hợp theo thời gian thực. Logstash có thể tự động thống nhất dữ liệu từ các nguồn khác nhau và chuẩn hóa dữ liệu theo ý của bạn.

- Logstash có 3 giai đoạn xử lí: inputs → filters → outputs.

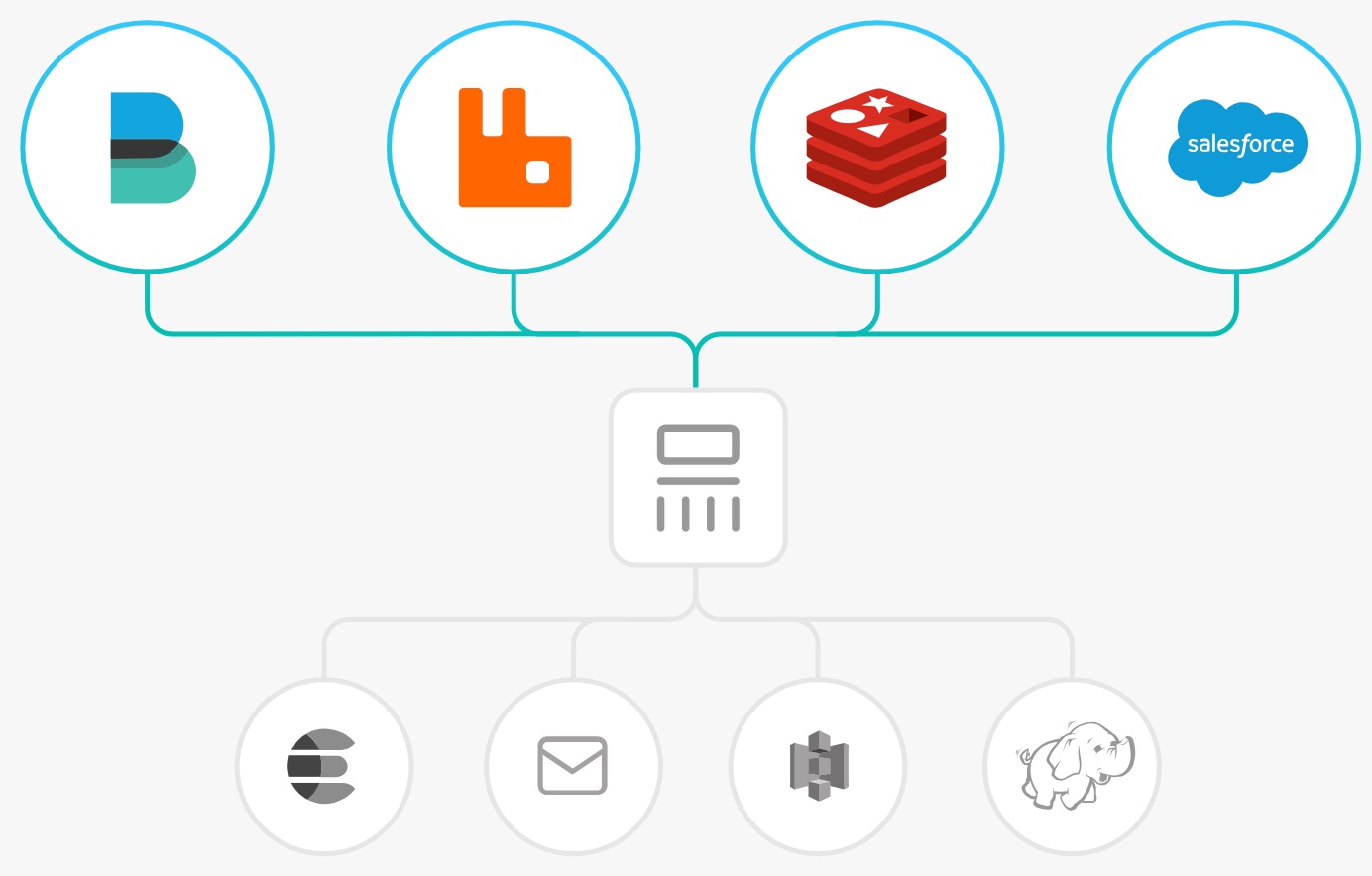


**2. Cách hoạt động**

**-** Đường ống xử lý sự kiện của Logstash có ba giai đoạn: input → filter → output. Các đầu vào tạo ra các sự kiện, bộ lọc sửa đổi chúng và các đầu ra sẽ chuyển chúng tới nơi khác. Đầu vào và đầu ra hỗ trợ codec cho phép bạn mã hóa hoặc giải mã dữ liệu khi nó vào hoặc thoát khỏi đường dẫn mà không phải sử dụng bộ lọc riêng biệt.

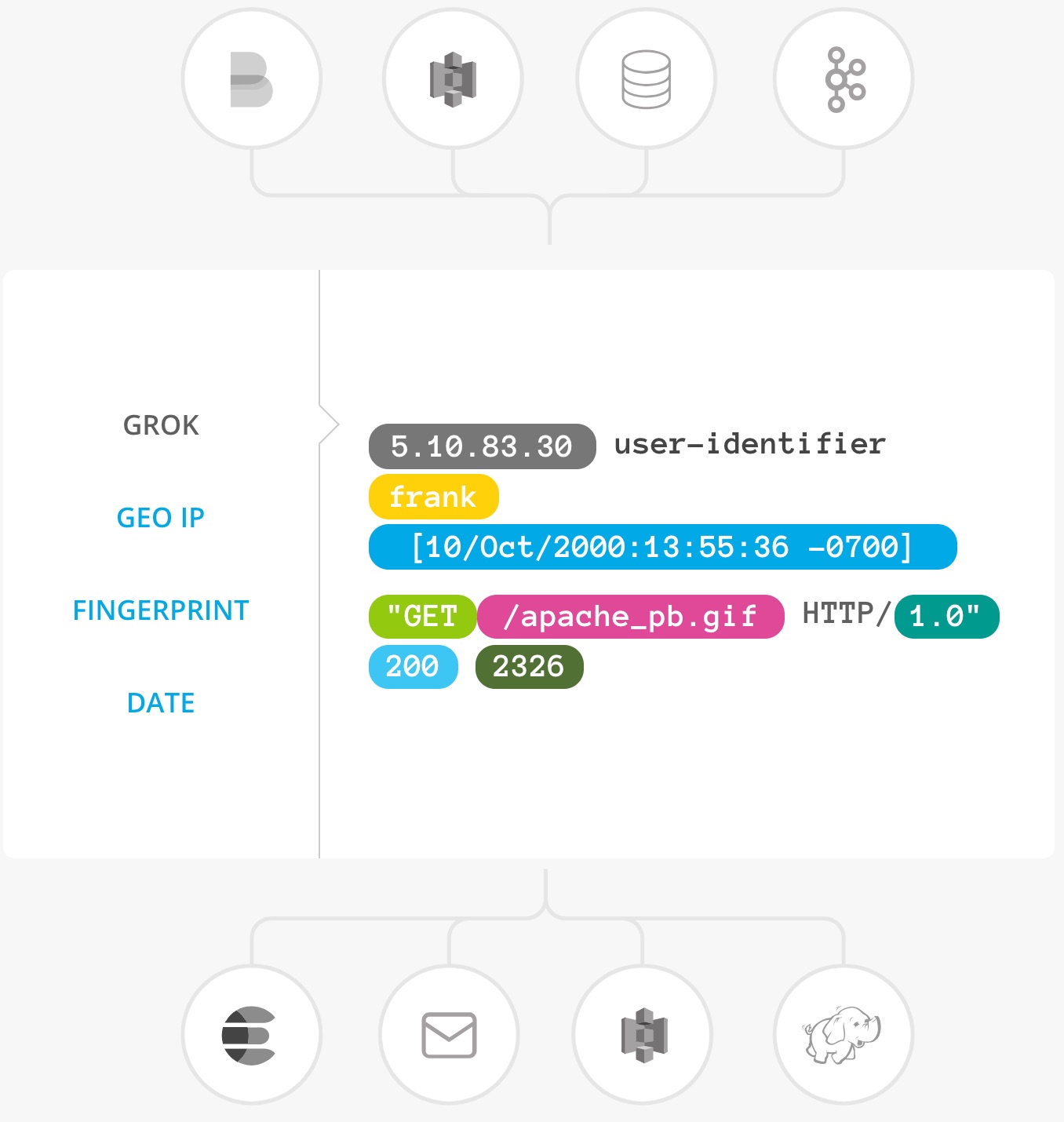
- Chúng ta sử dụng **Input** để lấy dữ liệu vào Logstash. Một số đầu vào thường được sử dụng là:

* File : đọc từ một tệp trên hệ thống
* Syslog : nghe trên cổng 514 cho các thông báo nhật ký hệ thống và format theo định dạng RFC3164.
* Beats : xử lý các sự kiện do beats gửi.



- Filter là thiết bị xử lý trung gian trong đường dẫn Logstash. Chúng ta có thể kết hợp các bộ lọc với các điều kiện để thực hiện một hành động trên một sự kiện nếu nó đáp ứng các tiêu chí nhất định. Một số bộ lọc hữu ích bao gồm :

* Grok : phân tích cú pháp và cấu trúc văn bản tùy ý - chỉnh sửa định dạng log từ client gửi về. Grok hiện là cách tốt nhất trong Logstash để phân tích cú pháp dữ liệu nhật ký không được cấu trúc thành một thứ có cấu trúc và có thể truy vấn được. Với 120 mẫu được tích hợp sẵn trong Logstash, nhiều khả năng chúng ta sẽ tìm thấy một mẫu đáp ứng nhu cầu của mình.
* Mutate : thực hiện các phép biến đổi chung trên các trường sự kiện. Bạn có thể đổi tên, xóa, thay thế và sửa đổi các trường trong sự kiện của mình.
* Drop : xóa hoàn toàn sự kiện, ví dụ: debug events.
* Clone : tạo bản sao của sự kiện, có thể thêm hoặc xóa các trường.
* Geoip : thêm thông tin về vị trí địa lý của địa chỉ IP (cũng hiển thị biểu đồ tuyệt vời trong Kibana).



- **Output** là bước cuối cùng trong chuỗi các bước xử lý của Logstash. Một sự kiện có thể đưa qua nhiều output khác nhau, tiếp đây là các **Output plugin** hay sử dụng.

* **elasticsearch**: gửi dữ liệu sự kiện đến hệ thống Elasticsearch. Tất nhiên đầu cuối của hệ thống logging ELK thường là Elasticsearch giúp bạn lưu trữ log, tìm kiếm log, …
* **file**: nếu bạn chả cần bất kì sự lưu trữ log cho việc tìm kiếm, hiển thị,… thì có thể lưu ra file trên hệ thống.
* **graphite**: gửi dữ liệu tới graphite, một trong những tool mã nguồn mở hỗ trợ việc lưu trữ và tạo biểu đồ metric.