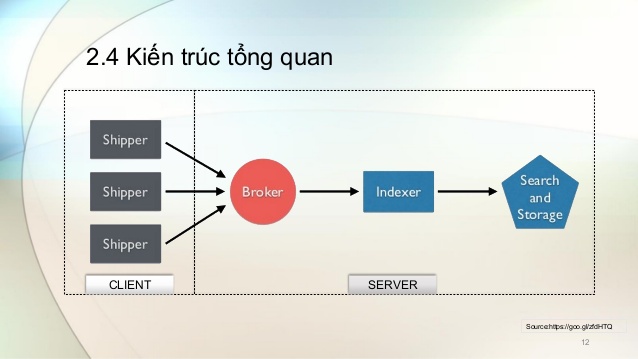
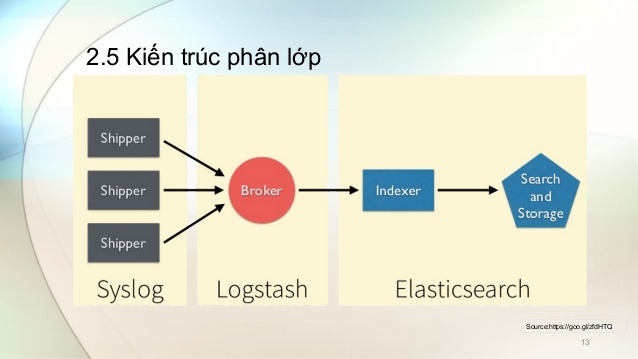
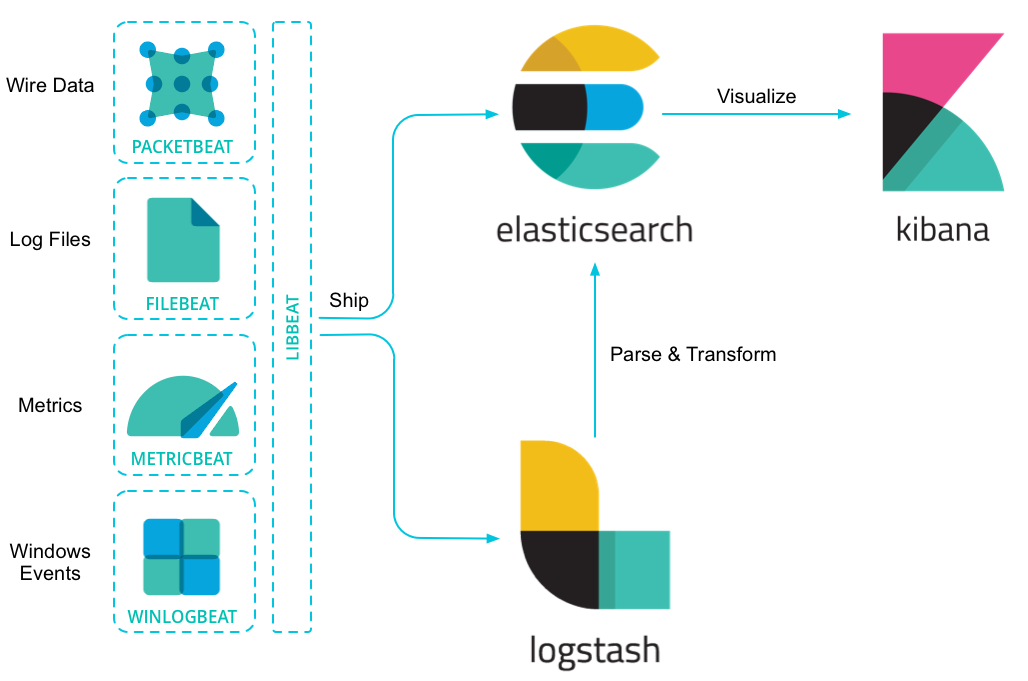
**Shipper – Broker - Indexer**







**1. Tổng quan**

Dữ liệu được các Beats (shipper) thu thập và gửi về cho Logstash; Logstash tiếp nhận và phân tích dữ liệu. Sau đó dữ liệu được gửi vào Elasticsearch; Elasticsearch nhận dữ liệu từ Logstash và lưu trữ, đánh chỉ mục; Kibana sử dụng các dữ liệu trong Elasticsearch để hiển thị và phân tích cú pháp tìm kiếm mà người dùng nhập vào để gửi cho Elasticsearch tìm kiếm.

Beats --> Logstash --> Elasticsearch <--> Kibana

**2. Shipper: gửi dữ liệu từ client tới ELK server**

**Sử dụng Beats**

Được viết bằng Golang, chạy trên các client để thu thập dữ liệu. Bạn có thể sử dụng libbeat để viết các beats cho mục đích thu thập của riêng mình.

Beat được cài đặt trên client với nhiệm vụ gửi dữ liệu về logstash

Hiện tại có các Beats được cung cấp sẵn bởi elastic là:

* Filebeat: đọc file và lưu vị trí cuối cùng, khi có dữ liệu mới sẽ đọc tiếp và gửi
* Packetbeat: capture gói tin trên các port của client, chuyển tiếp dữ liệu về Logstash

Các shipper gửi dữ liệu về Logstash, không đi qua bộ lọc nào nên được Logstash chuyển thành JSON và push thẳng sang Elasticsearch.

* Topbeat (metricbeat): Thu thập dữ liệu về tài nguyên hệ thống client và gửi về Logstash

Topbeat là một trong các "Beats" data shipper, thực hiện việc gửi các loại dữ liệu các nhau từ Server tới Elasticsearch. Cho phép bạn thu thập thông tin về CPU, memory, process activity.

* Winlobeat: Thu thập dữ liệu trên windows (có thể sử dụng nxlog để thu thập)

ELK cần sử dụng các "beat" để làm shipper giúp gửi các loại dữ liệu từ client tới Server.

Các beat index pattern cần được cài đặt trên cả ELK server và các client. Trên ELK server, các beat sẽ kết hợp với các thành phần để lọc dữ liệu, đánh chỉ mục, hiển thị.

**Các index pattern thường dùng**

* [packetbeat-]YYYY.MM.DD : thực hiện gửi dữ liệu capture từ các port về server
* [topbeat-]YYYY.MM.DD :
* [filebeat-]YYYY.MM.DD : thực hiện load file, đọc và gửi các dữ liệu mới về server
* [winlogbeat-]YYYY.MM.DD

**3. Broker – Logstash: có nhiệm vụ làm nơi tập trung log và phân tích dữ liệu**

Logstash là một công cụ mạnh mẽ cho việc nhận tập trung và phân tích log, giúp cho bạn có một cái nhìn về môi trường và hoạt động của server.

Nhận dữ liệu từ các beats, tiến hành phân tích dữ liệu:

* Phân tích dữ liệu gửi từ filebeat bằng grok
* Grok là một dạng khai báo pattern sử dụng regular expression
* Grok đã được khai báo rất nhiều pattern có sẵn, bạn có thể sử dụng ngay.

Nếu bạn có các loại dữ liệu đặc thù thì có thể dựa vào regular expression để khai báo các pattern theo yêu cầu

Các dữ liệu packetbeat, topbeat thì có các template sẵn nên không cần khai báo filter, được logstach gửi thẳng vào Elasticsearch

Dữ liệu được gửi sang cho Elasticsearch để lưu trữ.

Logstash có chức năng phân tích cú pháp của các dòng dữ liệu. Việc phân tích làm cho dữ liệu đầu vào ở một dạng khó đọc, chưa có nhãn thành một dạng dữ liệu có cấu trúc, được gán nhãn.

Khi cấu hình Logstash luôn có 3 phần: Input, Filter, Output.

Bình thường khi làm việc với Logstash, sẽ phải làm việc với Filter nhiều nhất. Filter hiện tại sử dụng Grok để phân tích dữ liệu

**4. Indexer – Elasticsearch: đánh chỉ mục dữ liệu.**

Elasticsearch có độ sẵn sàng cao và engine tìm kiếm phân tán. Mỗi index được cắt thành shard, mỗi shard có một hoặc nhiều replica. Mặc định, một index được tạo với 5 shard và 1 replica với mỗi shard (5/1). Có nhiều kỹ thuật có thể sử dụng, gồm 1/10 (tăng tốc độ tìm kiếm), hay 20/1 (tăng tốc độ đánh chỉ mục, việc tìm kiếm thực hiện trong một shard)