Log Management

jnlin

國立陽明交通大學資工系資訊中心

Computer Center of Department of Computer Science, NYCU

Log Management

- A system for gathering, processing and storing large volumes of logs, which were generated by operating systems, network appliance (switches, routers), or software applications.
    - Provides an interface for human reading, and may have APIs for program processing

Why Log Management (1)

- For Debug
  - Some bugs occurs only in particular situation, not always happen
  - ”Replay” the actions to reproduce the bug
  - “Test” new features without affecting customers
- For Audit
  - Who did the management? Who did “rm –rf” ?

Why Log Management (2)

- For Monitoring
  - Statistics from logs (e.g. HTTP 500s)
  - Abnormal numbers (increasement or decrement) means some parts of systems going wrong
- For AIOps
  - Help administrators to predict accidents by machine learning models
  - Data is the key part of AI

Log Management Key Points

- Accessible
  - Visual Dashboard
  - Row logs for deep debugging
- Durable
  - Can not be deleted / modified by anyone
  - Keep for reasonable time
    - Financial records: 7 years
- Realtime / near-realtime
  - Find accidents ASAP
  - Alert administrators if there is bad smell from statistics

Log Management Features

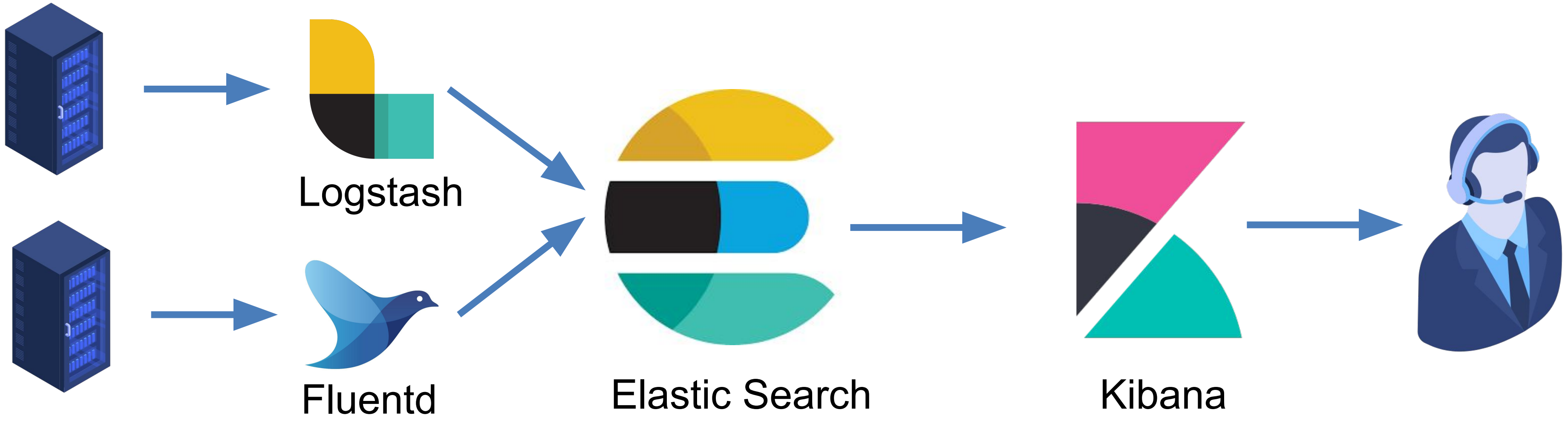
- Log collection
- Centralized log aggregation
- Long-term log storage and retention
- Log rotation
- Log analytics
- Log search and reporting

Log Management Software

- ELK / EFK
  - Elastic Search + Logstash (Fluentd) + Kibana
- Sentry (for application logs)
- Splunk (Commercial Software)
- Datadog (SAAS)
- Google Cloud Logging (Stackdriver)
- AWS Cloud Watch

ELK / EFK

- Elastic Search + Logstash (Fluentd) + Kibana
- Elastic Search
  - Storage and Index service for logs
- Logstash / Fluentd
  - Log collection, preprocessing and aggregation
- Kibana
  - Visual dashboard for log analytics



Elastic Search

- An open source, full-text search engine based on Apache Lucene
- Features
  - Distributed
  - Multitenancy
    - Serving multiple types of documents in one Elastic Search cluster
  - Near real-time search
- Developed and maintained by Elastic NV
  - The “open source business model”

Elastic Search – Basic Concept (1)

- Document
  - The base unit of storage of Elastic Search
    - “Row” in RDBMS
  - JSON format
  - Unique ID (UID)
- Index
  - The logical partition of documents
    - “Table” in RDBMS
  - Store similar documents
  - We can have multiple indices in one Elastic Search cluster

Elastic Search – Basic Concept (2)

- Nodes
  - The service instance running Elastic Search
  - Types
    - Master node
      - Maintain cluster state
      - Distribute shards to data nodes
      - Create and delete indices
    - Data node
      - Hot & warm node
      - Cold node
    - Ingest node
      - Pre-processing pipelines

Elastic Search – Basic Concept (3)

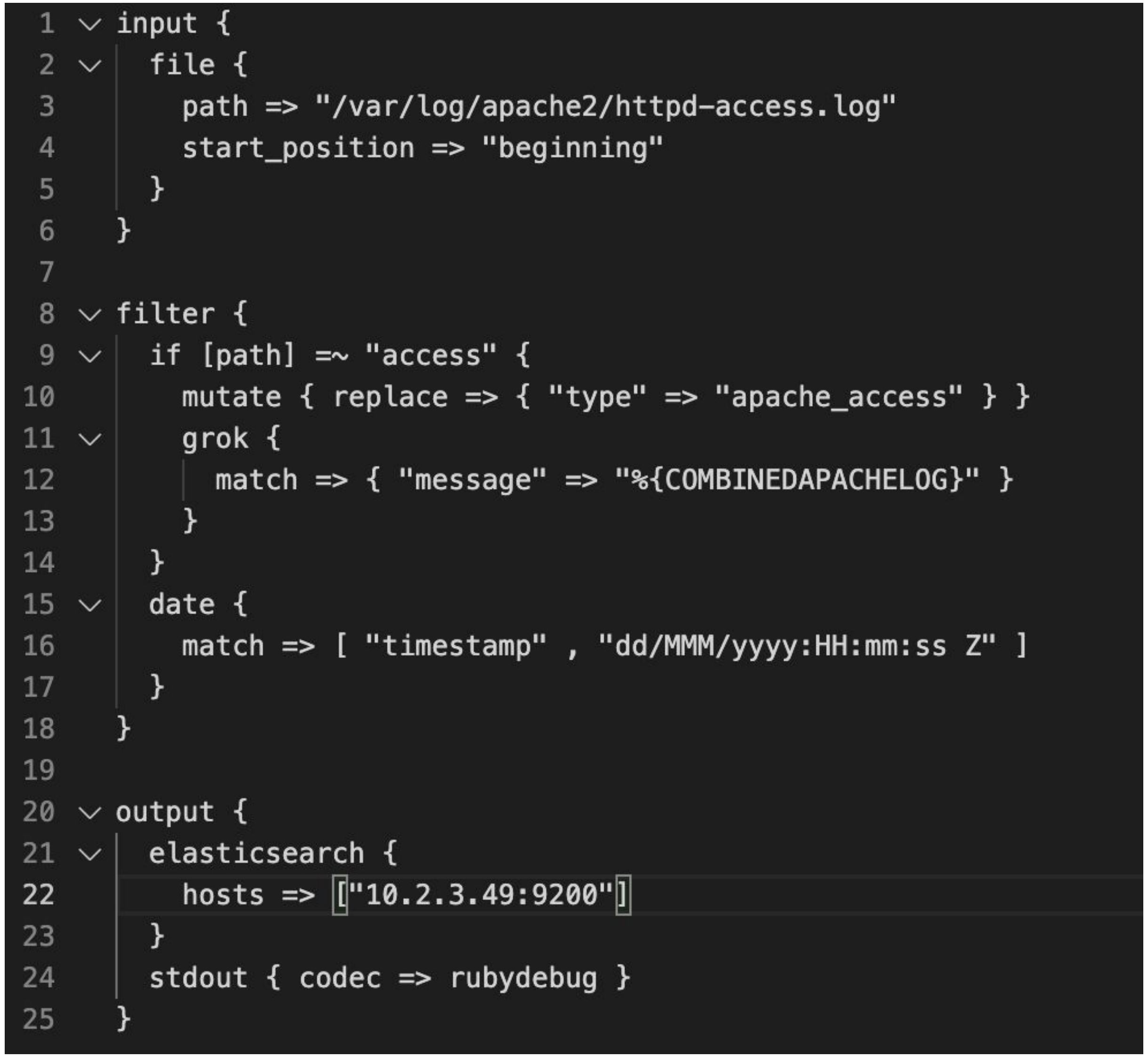
- Shards
  - Store index and documents
  - A single Lucene index
  - Indices will be split to serval shards
  - Shards will be duplicated for high availability

Logstash

- Collect, parse and transform logs
- An open source software developed by Elastic NV
- Support plugins for input, filtering and output
  - https://www.elastic.co/guide/en/logstash/current/input-plugins.html
  - https://www.elastic.co/guide/en/logstash/current/output-plugins.html
  - https://www.elastic.co/guide/en/logstash/current/filter-plugins.html
- Input
  - Web access logs / Syslogd / APIs / …
- Output
  - Elastic search/ IM (Slack / Discord) / Syslogd / …

Logstash – Configuration

- logstash.conf



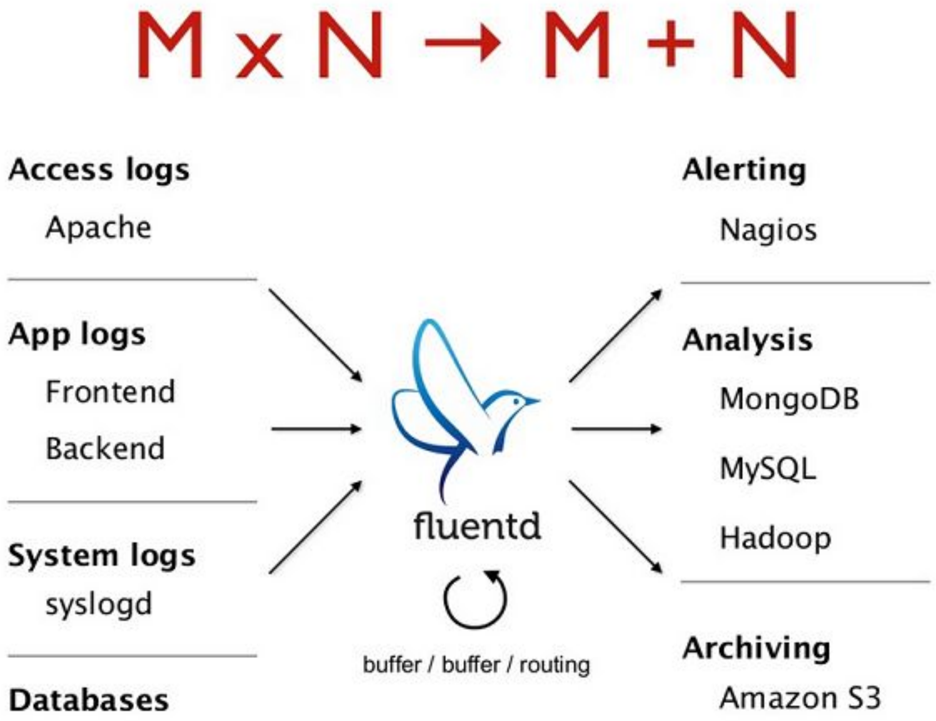
```
input {
  file {
    path => "/var/log/apache2/httpd-access.log"
    start_position => "beginning"
  }
}

filter {
  if [path] =~ "access" {
    mutate { replace => { "type" => "apache_access" } }
    grok {
      match => { "message" => "%{COMBINEDAPACHELOG}" }
    }
  }
  date {
    match => [ "timestamp" , "dd/MMM/yyyy:HH:mm:ss Z" ]
  }
}

output {
  elasticsearch {
    hosts => ["10.2.3.49:9200"]
  }
  stdout { codec => rubydebug }
}
```

Fluentd: Unified Logging Layer

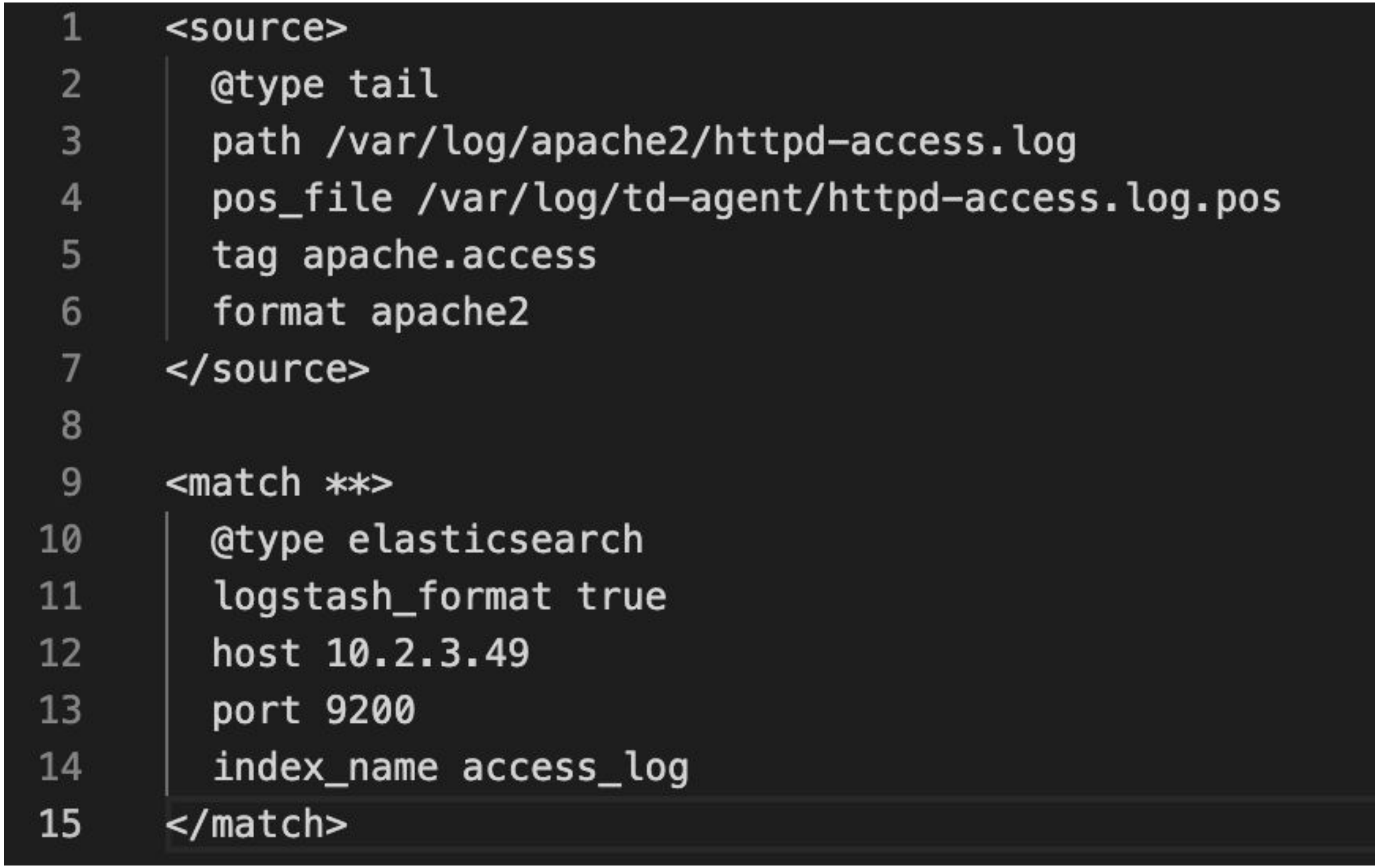
- An open source project for unifying the data collection and consumption
- Original developed by Treasure Data, now it is under the Cloud Native Computing Foundation (CNCF)
- https://www.fluentd.org
- Support plugins for data source, outputs and processing

Unified Logging Layer

- A layer for filtering, buffering and routing data
- Provides a unifying format (JSON) for data processing and transport
- Data buffering and retry-able data-transfer
- Horizontally scalable
- Reduce Complexity



Fluentd - Configuration



```
<source>
  @type tail
  path /var/log/apache2/httpd-access.log
  pos_file /var/log/td-agent/httpd-access.log.pos
  tag apache.access
  format apache2
</source>

<match **>
  @type elasticsearch
  logstash_format true
  host 10.2.3.49
  port 9200
  index_name access_log
</match>
```

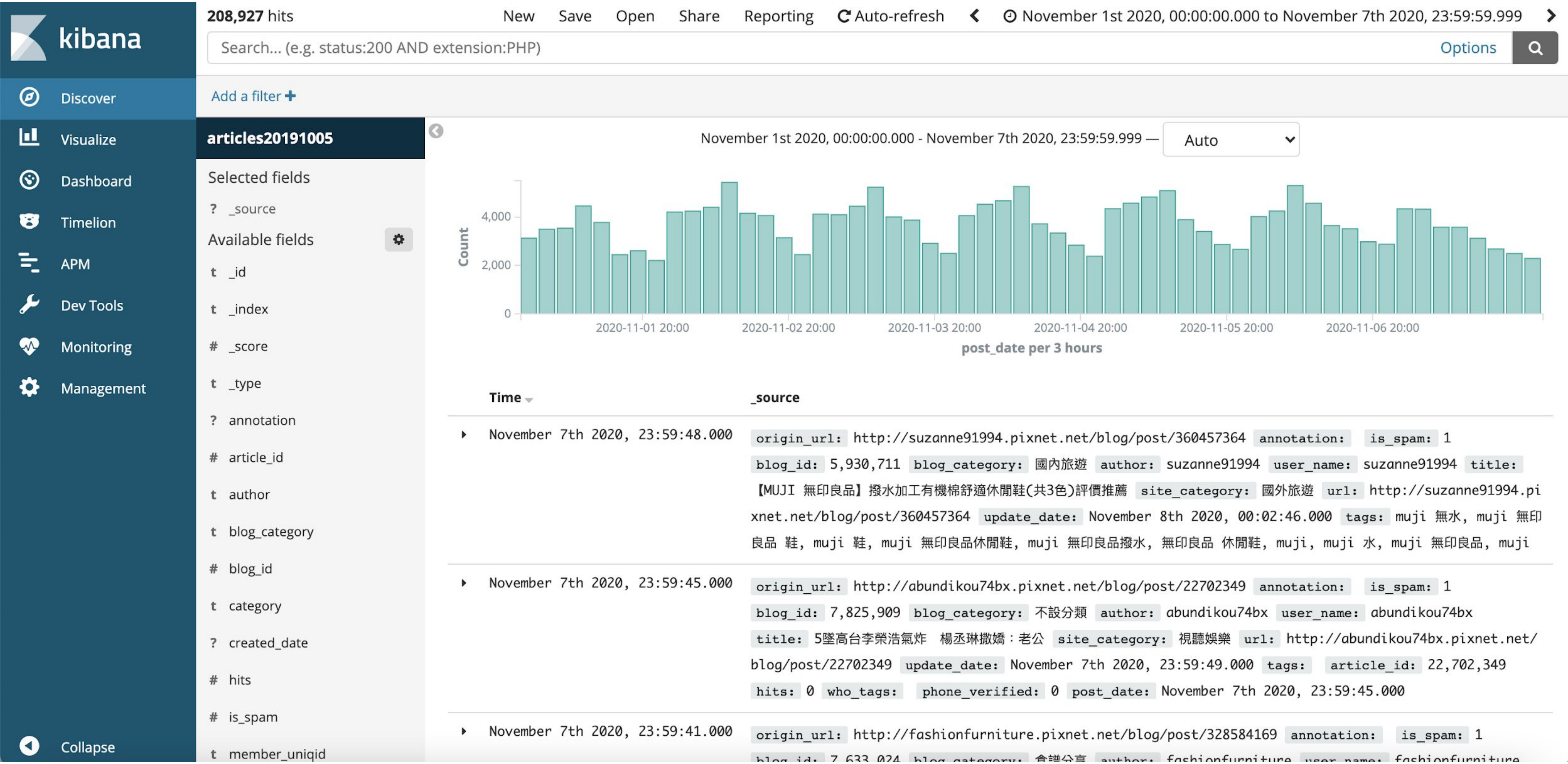
Beats

- A lightweight data collector, developed by Elastic NV
- Send collected data to Logstash or Elastic Search
- Centralized configuration in Kibana
- Types
  - Auditbeat: Audit data
  - Filebeat: Log files
  - Functionbeat: Cloud data
  - Heartbeat: Availability
  - Metricbeat: Metrics
  - Packetbeat: Network traffic
  - Winlogbeat: Windows event logs

Kibana

- Visual dashboard for users querying logs stored in Elastic Search
- An open source project developed by Elastic NV

Kibana - Dashboard



208,927 hits

November 1st 2020, 00:00:00.000 to November 7th 2020, 23:59:59.999

November 1st 2020, 00:00:00.000 - November 7th 2020, 23:59:59.999 —

```
November 7th 2020, 23:59:48.000   origin_url: http://suzanne91994.pixnet.net/blog/post/360457364 annotation: is_spam: 1 blog_id: 5,930,711 blog_category: 國內旅遊 author: suzanne91994 user_name: suzanne91994 title: 【MUJI 無印良品】撥水加工有機棉舒適休閒鞋(共3色)評價推薦 site_category: 國外旅遊 url: http://suzanne91994.pixnet.net/blog/post/360457364 update_date: November 8th 2020, 00:02:46.000 tags: muji 無水, muji 無印良品 鞋, muji 鞋, muji 無印良品休閒鞋, muji 無印良品撥水, 無印良品 休閒鞋, muji, muji 水, muji 無印良品, muji

November 7th 2020, 23:59:45.000   origin_url: http://abundikou74bx.pixnet.net/blog/post/22702349 annotation: is_spam: 1 blog_id: 7,825,909 blog_category: 不設分類 author: abundikou74bx user_name: abundikou74bx title: 5墜高台李榮浩氣炸　楊丞琳撒嬌：老公 site_category: 視聽娛樂 url: http://abundikou74bx.pixnet.net/blog/post/22702349 update_date: November 7th 2020, 23:59:49.000 tags: article_id: 22,702,349 hits: 0 who_tags: phone_verified: 0 post_date: November 7th 2020, 23:59:45.000

November 7th 2020, 23:59:41.000   origin_url: http://fashionfurniture.pixnet.net/blog/post/328584169 annotation: is_spam: 1 blog_id: 7,633,024 blog_category: 食譜分享 author: fashionfurniture user_name: fashionfurniture
```

Reference

- https://www.elastic.co/guide/en/elasticsearch/reference/current/getting-started.html
- https://www.elastic.co/guide/en/logstash/current/index.html
- https://docs.fluentd.org/