

Log Management

wangth

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

Log Management Features

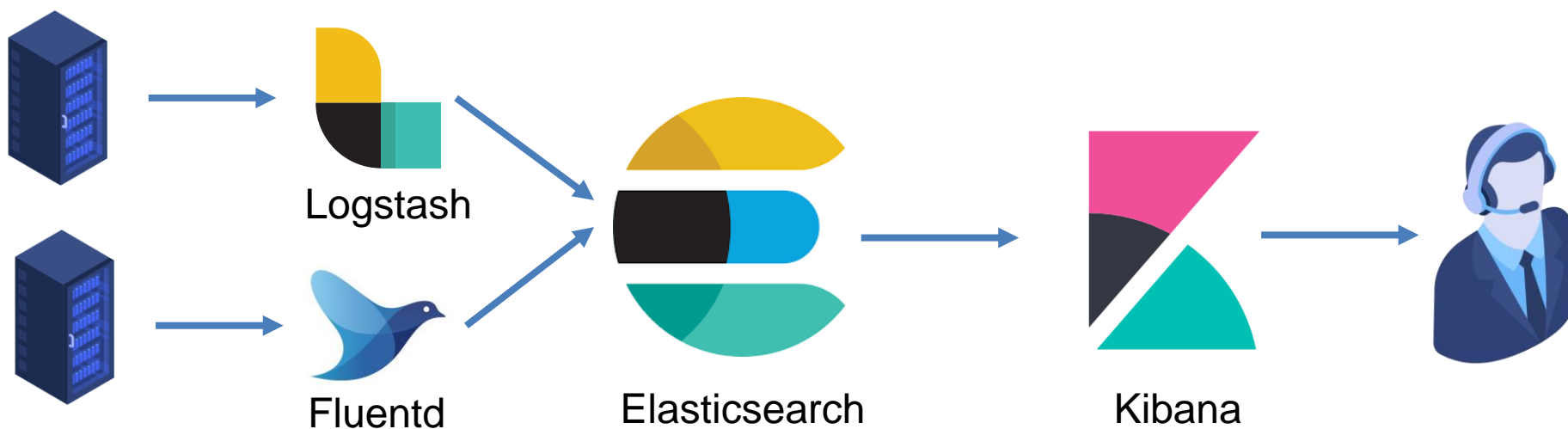
- Log collection
- Centralized log aggregation
- Long-term log storage and retention
- Log rotation
- Log analytics
- Log search and reporting

Log Management Software

- ELK / EFK
 - Elasticsearch + Logstash (Fluentd) + Kibana
- Sentry (Application Monitoring)
- Splunk (Data Platform)
- Datadog (SaaS, Cloud Monitoring as a Service)
- Google Cloud Logging (Stackdriver)
- AWS CloudWatch

ELK / EFK

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



Elasticsearch

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

Elasticsearch – Basic Concept (1)

- Document
 - The base unit of storage of Elasticsearch
 - “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 Elasticsearch cluster

Elasticsearch – Basic Concept (2)

- Nodes
 - The service instance running Elasticsearch
 - 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

Elasticsearch – Basic Concept (3)

- Shards
 - Store index and documents
 - A single Lucene index
 - Indices will be split to several 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
 - Elasticsearch/ IM (Slack / Discord) / Syslogd / ...

Logstash – Configuration

- logstash.conf

```
1  √ input {
2  √   file {
3      path => "/var/log/apache2/httpd-access.log"
4      start_position => "beginning"
5   }
6 }
7
8  √ filter {
9  √   if [path] =~ "access" {
10      mutate { replace => { "type" => "apache_access" } }
11  √   grok {
12      | match => { "message" => "%{COMBINEDAPACHELOG}" }
13   }
14 }
15  √ date {
16      match => [ "timestamp" , "dd/MMM/yyyy:HH:mm:ss Z" ]
17 }
18 }
19
20  √ output {
21  √   elasticsearch {
22      hosts => ["10.2.3.49:9200"]
23   }
24   stdout { codec => rubydebug }
25 }
```

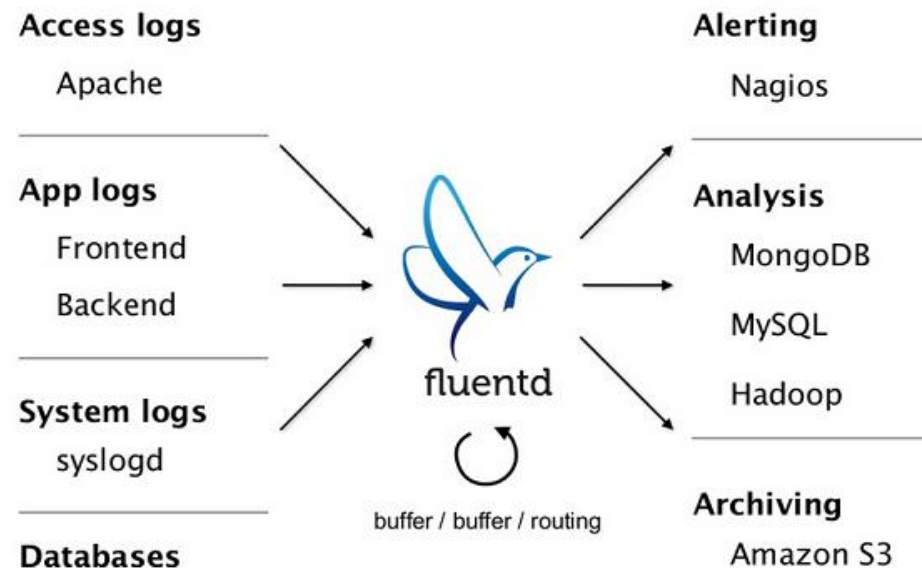
Fluentd: Unified Logging Layer

- 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

$$M \times N \rightarrow M + N$$



Fluentd - Configuration

```
1  <source>
2    @type tail
3    path /var/log/apache2/httpd-access.log
4    pos_file /var/log/td-agent/httpd-access.log.pos
5    tag apache.access
6    format apache2
7  </source>
8
9  <match **>
10   @type elasticsearch
11   logstash_format true
12   host 10.2.3.49
13   port 9200
14   index_name access_log
15 </match>
```

Beats

- A lightweight data collector, developed by Elastic NV
- Send collected data to Logstash or Elasticsearch
- 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 Elasticsearch
- An open-source project developed by Elastic NV

Kibana - Dashboard

kibana

- Discover
- Visualize
- Dashboard
- Timelion
- APM
- Dev Tools
- Monitoring
- Management

articles20191005

Selected fields

- ? `_source`

Available fields

- t `_id`
- t `_index`
- # `_score`
- t `_type`
- ? `annotation`
- # `article_id`
- t `author`
- t `blog_category`
- # `blog_id`
- t `category`
- ? `created_date`
- # `hits`
- # `is_spam`
- t `member_uniqid`

Collapse

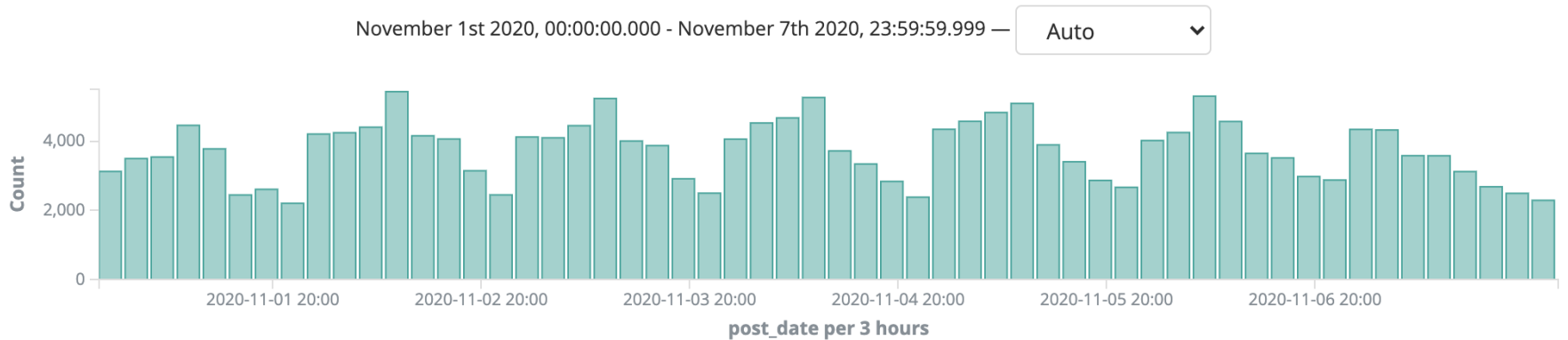
208,927 hits

New Save Open Share Reporting Auto-refresh November 1st 2020, 00:00:00.000 to November 7th 2020, 23:59:59.999

Search... (e.g. status:200 AND extension:PHP)

Options

Add a filter +



Time	_source
▶ November 7th 2020, 23:59:48.000	<code>origin_url</code> : http://suzanne91994.pixnet.net/blog/post/360457364 <code>annotation</code> : <code>is_spam</code> : 1 <code>blog_id</code> : 5,930,711 <code>blog_category</code> : 國內旅遊 <code>author</code> : suzanne91994 <code>user_name</code> : suzanne91994 <code>title</code> : 【MUJI 無印良品】撥水加工有機棉舒適休閒鞋(共3色)評價推薦 <code>site_category</code> : 國外旅遊 <code>url</code> : http://suzanne91994.pixnet.net/blog/post/360457364 <code>update_date</code> : November 8th 2020, 00:02:46.000 <code>tags</code> : 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	<code>origin_url</code> : http://abundikou74bx.pixnet.net/blog/post/22702349 <code>annotation</code> : <code>is_spam</code> : 1 <code>blog_id</code> : 7,825,909 <code>blog_category</code> : 不設分類 <code>author</code> : abundikou74bx <code>user_name</code> : abundikou74bx <code>title</code> : 5墜高台李榮浩氣炸 楊丞琳撒嬌：老公 <code>site_category</code> : 視聽娛樂 <code>url</code> : http://abundikou74bx.pixnet.net/blog/post/22702349 <code>update_date</code> : November 7th 2020, 23:59:49.000 <code>tags</code> : <code>article_id</code> : 22,702,349 <code>hits</code> : 0 <code>who_tags</code> : <code>phone_verified</code> : 0 <code>post_date</code> : November 7th 2020, 23:59:45.000
▶ November 7th 2020, 23:59:41.000	<code>origin_url</code> : http://fashionfurniture.pixnet.net/blog/post/328584169 <code>annotation</code> : <code>is_spam</code> : 1 <code>blog_id</code> : 7,633,024 <code>blog_category</code> : 食譜分享 <code>author</code> : fashionfurniture <code>user_name</code> : 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/>