elastic | 中国开发者大会

Elastic Stack 在日志实践中的全链路监控实践

魏彬
蒋云结

个人简介

上海普翔

技术顾问

Elastic Certified Engineer

魏彬

rockybean

蒋云结

Leon J

elastic

运维工程师小明与 ELK 的故事

1年前的某天

BOSS：

小明!!! 用户投诉网站访问有问题，出什么问题了？

小明：

收到！
（OMG！20台 Web 服务器，要一台台登录查看）

1个小时过去了。。。。。。

BOSS：

小明，定位到问题了吗？怎么这么慢？！

小明：

快了，快了……

（OMG！还有5台日志没查，前面15台没有问题啊！）

最后在 **第19台** 服务器的 web server 日志中找到了大量报错信息，经过开发团队确认后，发现是这台服务器 **代码发布失败** 导致的！

小明的独白

长此以往，必成背锅侠！

有没有办法不登录服务器就可以查看日志呢？

Google [日志收集解决方案]：

分布式实时日志分析解决方案ELK部署架构- FEINIK的个人主页
https://my.oschina.net/feinik/blog/1580625 ▾ Translate this page
Jump to 2.2、Filebeat作为日志收集器 - 该架构与第一种架构唯一不同的是：应用端日志收集器换成了Filebeat，Filebeat轻量，占用服务器资源少，所以 ...

构建日志分析解决方案 - Amazon AWS
https://aws.amazon.com/cn/getting.../build-log-analytics-solution/ ▾ Translate this page
在本项目中，您将使用Amazon Web Services 来构建一个端到端日志分析解决方案，从而收集、提取、处理及加载批量数据和流数据，并让用户在其已有分析系统中 ...

集中式日志解决方案- 简书
https://www.jianshu.com/p/e3ccb75bd813 ▾ Translate this page
Sep 18, 2017 - 因此，我们需要一种数据收集框架，它可以将不同服务器上的日志数据， ... Trend的信息显示，ELK已经成为目前最流行的的集中式日志解决方案。

日志收集系统解决方案探讨- 天的CSDN - CSDN博客
https://blog.csdn.net/dutianmin/article/details/8898763 ▾ Translate this page
May 8, 2013 - 日志收集系统解决方案探讨 dutianmin@gmail..com 一、 需求背景 公司许多平台系统每天都会产生很多日志，日志的产生是源源不断的流数据（如 ...

分布式实时日志分析解决方案ELK 部署架构- 云+社区- 腾讯云
https://cloud.tencent.com/developer/news/283119 ▾ Translate this page
Jul 24, 2018 - ELK 已经成为目前最流行的集中式日志解决方案，它主要是由Beats、Logstash、Elasticsearch、Kibana等组件组成，来共同完成实时日志的收集， ...

集中式日志收集概述– Tiantian Gao (gtt116)
https://www.gaott.info/centralized-logging-system/ ▾ Translate this page

ELK!

ELK!!

ELK!!!

ELK

Kibana

Elasticsearch

Logstash

elastic

ELK -> ElasticStack

Kibana

Elasticsearch

Beats

Logstash

elastic

小明的独白

就它了！

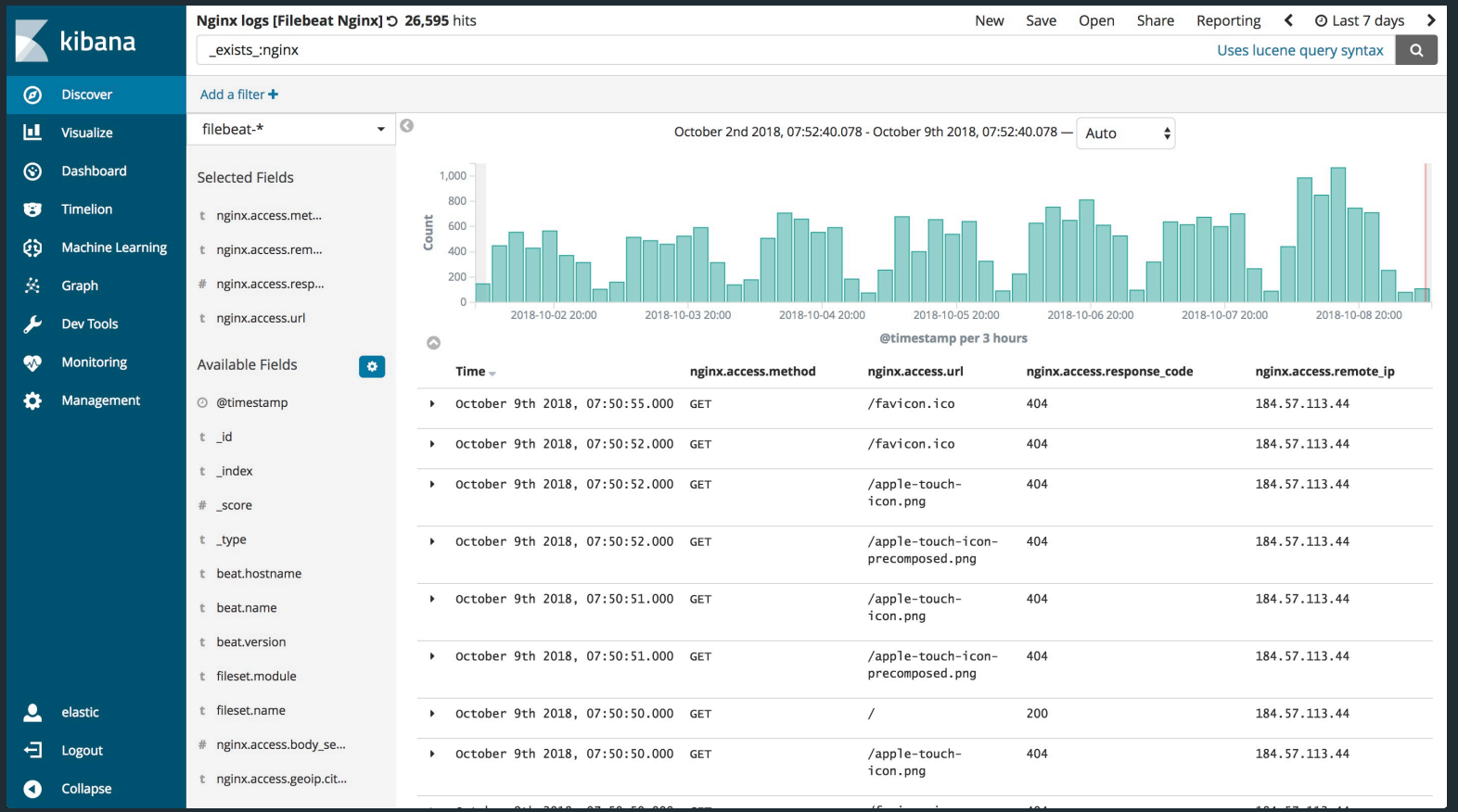
1天过去了。。。。。。

架构

简单　　快速　　有效

elastic

小明：


kibana
Nginx logs [Filebeat Nginx] 26,595 hits
New
Save
Open
Share
Reporting
Last 7 days
exists:nginx
Uses lucene query syntax
Discover
Visualize
Dashboard
Timelion
Machine Learning
Graph
Dev Tools
Monitoring
Management
elastic
Logout
Collapse
Add a filter +
filebeat-*
Selected Fields
t nginx.access.met...
t nginx.access.rem...
nginx.access.resp...
t nginx.access.url
Available Fields
@timestamp
t _id
t _index
_score
t _type
t beat.hostname
t beat.name
t beat.version
t fileset.module
t fileset.name
nginx.access.body_se...
t nginx.access.geoip.cit...
October 2nd 2018, 07:52:40.078 - October 9th 2018, 07:52:40.078 —
Auto
Count
1,000
800
600
400
200
0
2018-10-02 20:00
2018-10-03 20:00
2018-10-04 20:00
2018-10-05 20:00
2018-10-06 20:00
2018-10-07 20:00
2018-10-08 20:00
@timestamp per 3 hours
Time
nginx.access.method
nginx.access.url
nginx.access.response_code
nginx.access.remote_ip
October 9th 2018, 07:50:55.000 GET /favicon.ico 404 184.57.113.44
October 9th 2018, 07:50:52.000 GET /favicon.ico 404 184.57.113.44
October 9th 2018, 07:50:52.000 GET /apple-touch-icon.png 404 184.57.113.44
October 9th 2018, 07:50:52.000 GET /apple-touch-icon-precomposed.png 404 184.57.113.44
October 9th 2018, 07:50:51.000 GET /apple-touch-icon.png 404 184.57.113.44
October 9th 2018, 07:50:51.000 GET /apple-touch-icon-precomposed.png 404 184.57.113.44
October 9th 2018, 07:50:50.000 GET / 200 184.57.113.44
October 9th 2018, 07:50:50.000 GET /apple-touch-icon.png 404 184.57.113.44

小明的独白

棒极了！
终于可以不用登录服务器就可以查看日志了！

1周后。。。。。。

小明的独白

日志检索现在可以了，
能否分析下返回码 500 的请求量？

1天过去了。。。。。。

架构

elastic

半年时间过去了。。。

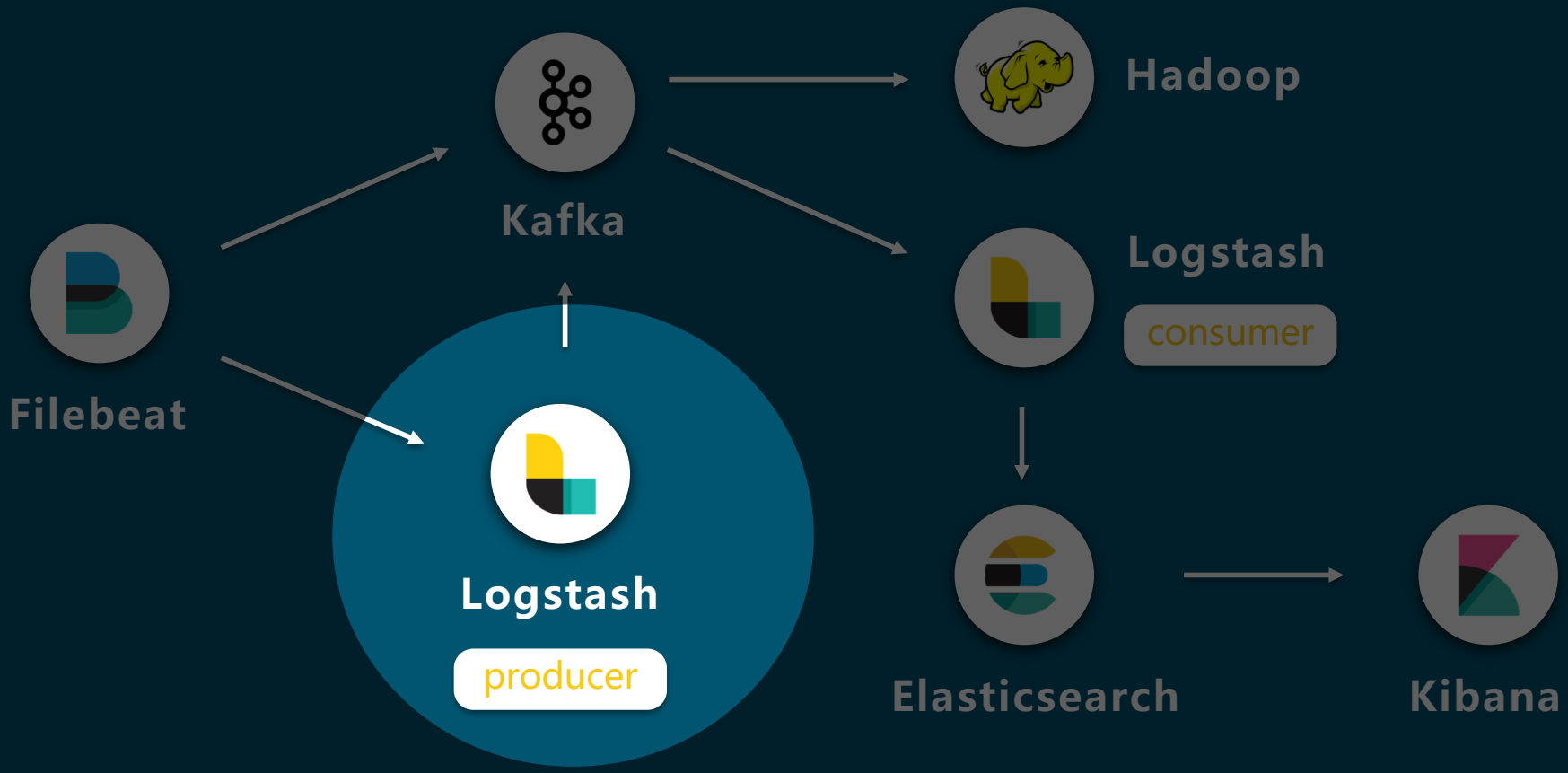
越来越多的日志和数据接入集群
小明也升职加薪为 ELK 集群运维负责人

小明接到新的需求

大数据团队希望可以
将数据导入其他大数据平台作分析，
这怎么解决？

三天过去了。。。

架构


Hadoop
Kafka
Logstash
consumer
Filebeat
Logstash
producer
Elasticsearch
Kibana

elastic

越来越多的团队使用 Kibana 来查询分析数据

1个月后的某天

团队 A：

小明，我们日志搜不到了，可服务器上明明有啊！

小明：

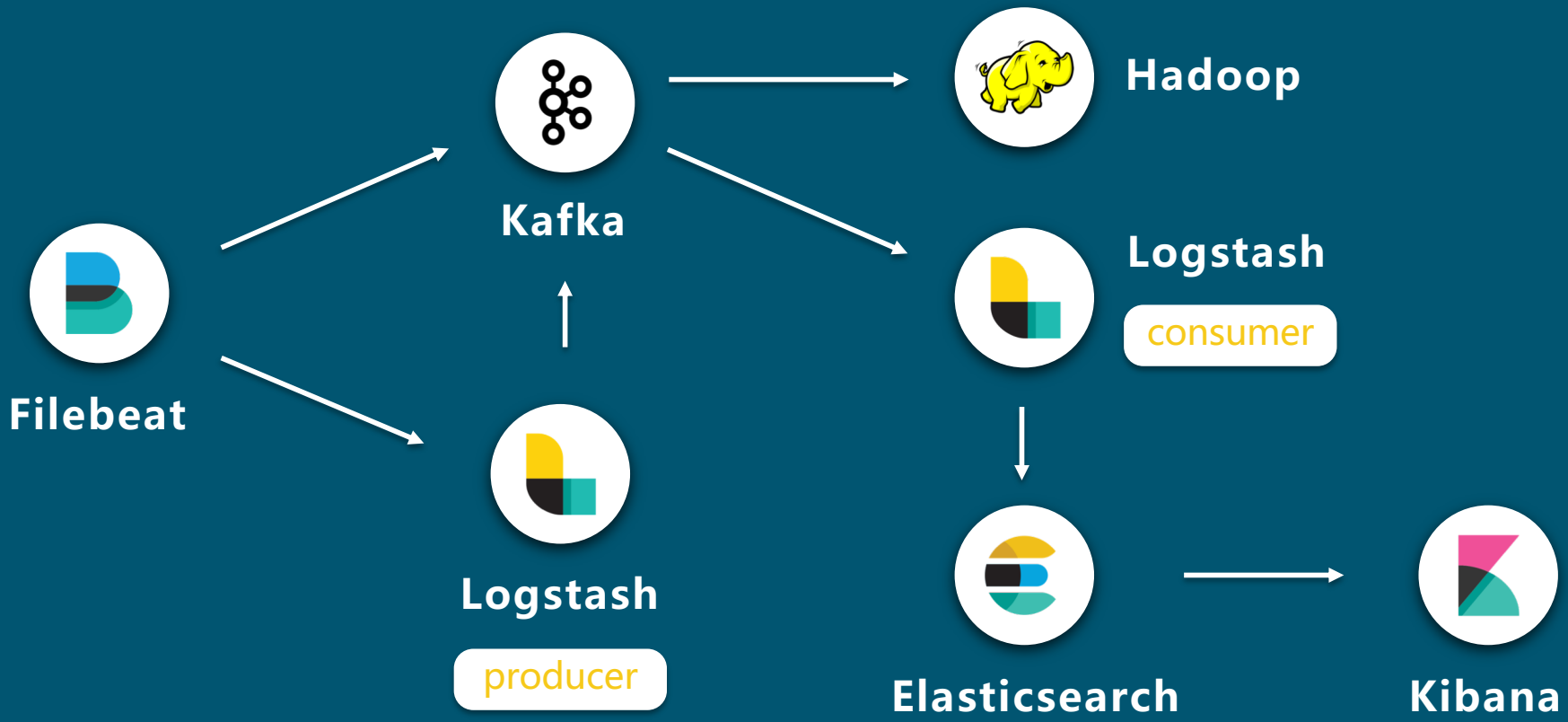
我来查下！

（OMG！他们在我这里存了多少日志？！这可怎么查？）

OMG！他们在我这里存了多少日志？！这可怎么查？

业务 A	业务 B	业务 C
5类日志 20台机器 3个 Index	2类日志 10台机器 2个 Index	10类日志 50台机器 6个 Index

架构


Hadoop
Kafka
Logstash
consumer
Filebeat
Logstash
producer
Elasticsearch
Kibana

elastic

如何排查？

第一步？

验证问题！！

哪个索引？哪台机器？哪个文件？

链路排查



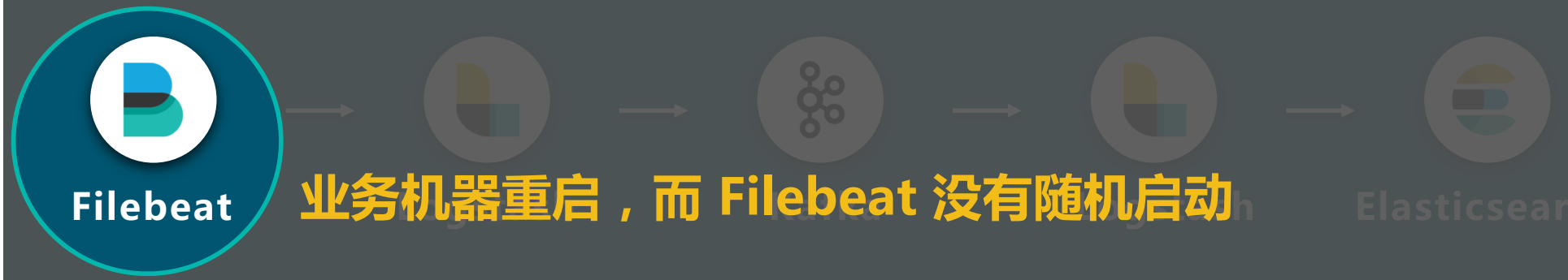


Filebeat	Logstash	Kafka	Logstash	Elasticsearch
Q: • FB进程还在吗？ • 日志有报错吗？ • FB 性能如何？ • 机器负载如何？ • 网络是否有问题？	Q: • LS进程还在吗？ • 日志有报错吗？ • LS 性能如何？ • 机器负载如何？ • 网络是否有问题？	Q: • Kafka 进程还在吗？ • 日志有报错吗？ • Kafka 性能如何？ • 机器负载如何？ • 网络是否有问题？	Q: • LS进程还在吗？ • 日志有报错吗？ • LS 性能如何？ • 机器负载如何？ • 网络是否有问题？	Q: • ES 集群状态？ • 日志有报错吗？ • ES 性能如何？ • 机器负载如何？ • 网络是否有问题？
A: • PS • cat log • Metric API • cpu mem disk • nettools	A: • PS • cat log • Metric API • cpu mem disk • nettools	A: • PS • cat log • Metric API • cpu mem disk • nettools	A: • PS • cat log • Metric API • cpu mem disk • nettools	A: • cat cluster health • cat log • Metric API • cpu mem disk • nettools

按照链路一层层排查问题。。。

半天过去了。。。

链路排查



业务机器重启，而 Filebeat 没有随机启动

由于某条日志过长，超过 Kafka 限制，
Logstash Producer 一直发送失败，导致
Logstash Consumer 解析报错，无法正常写入 ES

小明的独白

排查故障花了好长时间，
有没有高效的方法？！

链路排查





Filebeat	Logstash	Kafka	Logstash	Elasticsearch
Q: • FB进程还在吗？ • 日志有报错吗？ • FB 性能如何？ • 机器负载如何？ • 网络是否有问题？	Q: • LS进程还在吗？ • 日志有报错吗？ • LS 性能如何？ • 机器负载如何？ • 网络是否有问题？	Q: • Kafka 进程还在吗？ • 日志有报错吗？ • Kafka 性能如何？ • 机器负载如何？ • 网络是否有问题？	Q: • LS进程还在吗？ • 日志有报错吗？ • LS 性能如何？ • 机器负载如何？ • 网络是否有问题？	Q: • ES 集群状态？ • 日志有报错吗？ • ES 性能如何？ • 机器负载如何？ • 网络是否有问题？
A: • PS • cat log • Metric API • cpu mem disk • nettools	A: • PS • cat log • Metric API • cpu mem disk • nettools	A: • PS • cat log • Metric API • cpu mem disk • nettools	A: • PS • cat log • Metric API • cpu mem disk • nettools	A: • cat cluster health • cat log • Metric API • cpu mem disk • nettools

链路排查

Q:
- 进程还在吗？
- 日志有报错吗？
- 性能如何？
- 机器负载如何？
- 网络是否有问题？

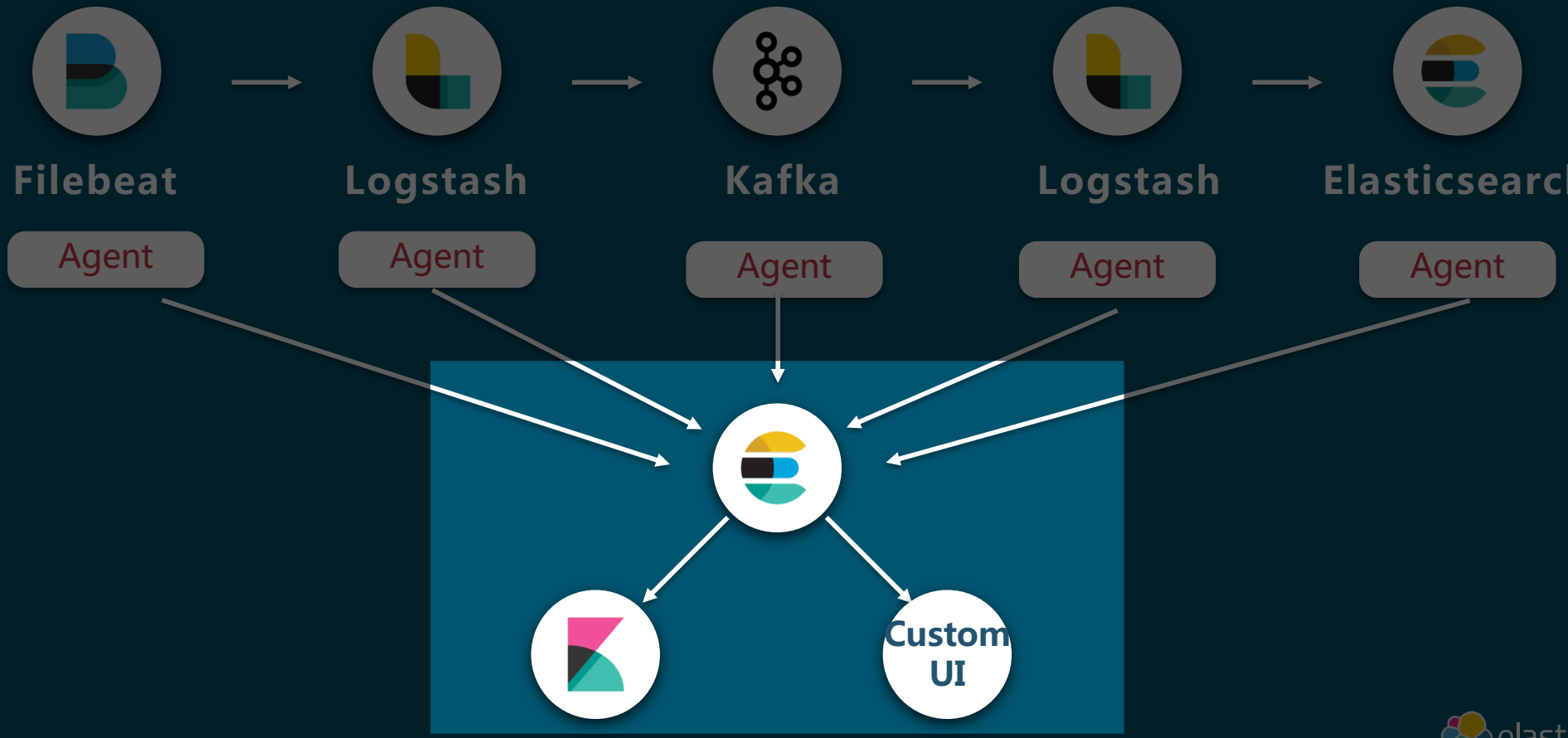
A:
- PS
- cat log
- Metric API
- cpu mem disk
- nettools

elastic

小明的独白

把每个节点的信息收集上来不就好了？！
这不就是 ElasticStack 最擅长的吗？

系统架构



elastic

Agent 作用

进程还在吗？	→	Heartbeat
日志有报错吗？	→	Filebeat
性能如何？	→	Metricbeat
机器负载如何？	→	Metricbeat
网络是否有问题？	→	Heartbeat

系统架构

日志产生

日志收集

延迟
Latency

elastic

小明的独白

有没有办法知道日志到达每个链路节点的时间？
这样就可以计算在每个路径的延迟了!!!

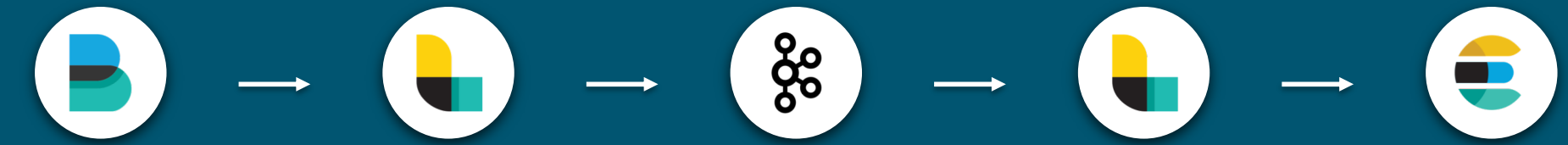
2018-11-10T08:00:00.0000+0800 INFO I'm logging

@timestamp
日志产生时间

```
{
  "@timestamp":"2018-11-10T08:00:00.0000+0800",
  "source":"/var/log/test.log",
  "message":"2018-11-10T08:00:00.0000+0800 INFO I'm logging"
}
```

延迟 Latency

延迟 Latency



@beat_ts

@ls_p_ts

@ls_c_ts

日志产生时间： @timestamp

Filebeat 采集延迟： @beat_ts - @timestamp

Logstash Producer 延迟： @ls_p_ts - @beat_ts

Logstash Consumer 延迟： @ls_c_ts - @ls_p_ts

elastic

Logstash 设置时间及计算 Latency

```
ruby {
        code => "event.set('@ls_c_ts', Time.now());"
}

ruby {
        code => "event.set('ls_c_latency',
[(event.get('@ls_c_ts').to_f - event.get('@ls_p_ts').to_f) *
1000, 0].max)"
}
```

```
{
  "@timestamp": "2018-11-10T08:00:00.000+0800",
  "@beat_ts": "2018-11-10T08:00:00.100+0800",
  "@ls_p_ts": "2018-11-10T08:00:01.030+0800",
  "@ls_c_ts": "2018-11-10T08:00:02.000+0800",
  "beat_latency": 100,
  "ls_p_latency": 930,
  "ls_c_latency": 970,
  "source": "/var/log/test.log",
  "message": "2018-11-10T08:00:00.000+0800 INFO I'm logging"
}
```

小明的独白

日志数据都收集上来了，排障速度杠杠的！
有没有办法实现自动化报警？！
不要每次等用户发现来抱怨？

告警机制

人工设置阈值

异常检测

X-Pack Machine Learning

告警机制

elastic

小明的独白

完美！

某天

BOSS：

小明，我们现在收集了多少业务的日志了？

小明：

我看看。。。

（OMG！日志收集的时候没记录业务啊……）

BOSS：

每个业务收集了多少种日志？他们都有多少机器啊？

小明：

不知道啊。。。

BOSS：

抓紧知道一下！

小明的独白

这该如何解决呢？

解决方案

采集处打上业务标签

fields.biz_name: CMS

ES 的聚合分析，即可快速展示业务

elastic

```
{
  "@timestamp": "2018-11-10T08:00:00.000+0800",
  "@beat_ts": "2018-11-10T08:00:00.100+0800",
  "@ls_p_ts": "2018-11-10T08:00:01.030+0800",
  "@ls_c_ts": "2018-11-10T08:00:02.000+0800",
  "beat_latency": 100,
  "ls_p_latency": 930,
  "ls_c_latency": 970,
  "fields":{
    "biz_name":"CMS"
  },
  "source": "/var/log/test.log",
  "message": "2018-11-10T08:00:00.000+0800 INFO I'm logging"
}
```

小明的独白

是不是也可以显示每个业务用到了哪些软件？
比如 nginx、redis、mysql 等
这样通过日志就可以反向推导出业务架构组成了

未完待续。。。

小明的Demo

elastic
中文社区

专业、垂直、纯粹的 Elastic 开源技术交流社区
https://elasticsearch.cn/

elastic