Elk的建设和微服务日志的收集

# 1 前言

Elastic Stack 提供 Beats 和 Logstash 套件来采集任何来源、任何格式的数据。其实Beats 和 Logstash的功能差不多，都能够与 Elasticsearch 产生协同作用，而且logstash比filebeat功能更强大一点，2个都使用是因为：Beats 是一个轻量级的采集器，支持从边缘机器向 Logstash 和 Elasticsearch 发送数据。考虑到 Logstash 占用系统资源较多，我们采用 Filebeat 来作为我们的日志采集器。并且这里采用kafka作为传输方式是为了避免堵塞和丢失，以实现日志的实时更新。

# 2 介绍

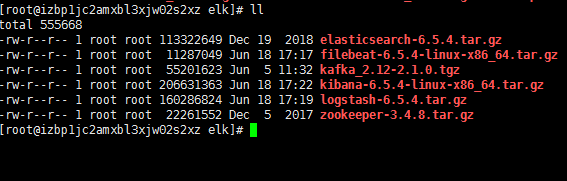
1、Filebeat：filebat是一个用于转发和集中日志数据的轻量级shipper。作为代理安装在服务器上，filebeat监视指定的日志文件或位置，收集日志事件，并将它们转发给ElasticSearch或logstash进行索引。

2、Logstash：Logstash 是开源的服务器端数据处理管道，能够同时从多个来源采集数据，转换数据，然后将数据发送到存储库。

3、ElasticSearch：Elasticsearch 是基于 JSON 的分布式搜索和分析引擎，专为实现水平扩展、高可靠性和管理便捷性而设计。

4、Kibana：Kibana 能够以图表的形式呈现数据，并且具有可扩展的用户界面，供全方位配置和管理 Elastic Stack。

# 3 ELK软件版本



kafka、Zookeeper安装

​ kafka是一种高吞吐量的分布式发布订阅消息系统，它可以处理消费者规模的网站中的所有动作流数据。

ZooKeeper是一个分布式的，开放源码的分布式应用程序协调服务，是Google的Chubby一个开源的实现，是Hadoop和Hbase的重要组件。它是一个为分布式应用

提供一致性服务的软件，提供的功能包括：配置维护、域名服务、分布式同步、组服务等。

# 4 Zookeeper 的安装

# 解压到 /usr/local/ 对应的目录下

tar -zxvf zookeeper-3.4.8.tar.gz -C /usr/local/

​

# 进入 zookeeper 目录

cd /usr/local/zookeeper-3.4.8/

​

# 修改配置 先复制模板配置文件，并重命名，然后里面存放数据的路径dataDir可以自己定义

cp conf/zoo\_sample.cfg conf/zoo.cfg

​

# 采用vim | vi 编辑文件

vim conf/zoo.cfg

​

# 启动服务

./bin/zkServer.sh start

​

# 查看是否启动成功

./bin/zkServer.sh status

​

# 查看 zookeeper

./bin/zkServer.sh status

ZooKeeper JMX enabled by default

Using config: /usr/local/zookeeper-3.4.8/bin/../conf/zoo.cfg

Mode: standalone

​

# Mode: standalone 说明启动成功

​

# 5 Kafa的安装

# 解压到 /usr/local/ 对应的目录下

tar -zxvf kafka\_2.12-2.1.0.tgz -C /usr/local/

​

# 进入 kafka 目录

cd /usr/local/kafka\_2.12-2.1.0/

​

# 修改配置文件

vim config/server.properties

​

# 启动kafka

./bin/kafka-server-start.sh config/server.properties

​

# 查看java进程pid   jps -l

jps -l

4512 org.apache.zookeeper.server.quorum.QuorumPeerMain

15684 org.logstash.Logstash

10069 kafka.Kafka

16317 sun.tools.jps.Jps

13678 org.elasticsearch.bootstrap.Elasticsearch

​

# Create a topic

./bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic test

​

# 查看 topic

./bin/kafka-topics.sh --list --zookeeper localhost:2181

​

# 查询 topic 详细信息

./bin/kafka-topics.sh --describe --zookeeper localhost:2181 --topic test

​

# ​6 FileBeat 安装

Filebeat是一个轻量级的数据传输者，处理流程基本分为3部分：Input、Filter、Output。

# 解压到 /usr/local/ 对应的目录下

tar -zxvf filebeat-6.5.4-linux-x86\_64.tar.gz -C /usr/local/

​

# 进入 filebeat 目录 修改filebeat.yml，filebeat默认是将数据传输到elasticsearch，需要修改为kafka

cd /usr/local/filebeat-6.5.4/

​

# 修改配置文件

vim filebeat.yml

​

# 启动 filebeat

sudo ./filebeat -e -c filebeat.yml

无误后

nohup ./filebeat -e -c filebeat.yml >&/dev/null &

​

# 6 elasticlsearch

Elasticsearch 是一个分布式、RESTful 风格的搜索和数据分析引擎，能够解决不断涌现出的各种用例。

# 解压到 /usr/local/ 对应的目录下

tar -zvxf elasticsearch-6.5.4.tar.gz -C /usr/local/

​

# 进入 elasticsearch 目录

cd /usr/local/elasticsearch-6.5.4/

​

# 注：安装 elasticsearch 这里太多坑 接下来一步一步细说

​

# 1、第一步修改 config/elasticsearch.yml 配置文件

vim config/elasticsearch.yml

​

# 2、新增用户 es，因为默认情况下 elasticsearch 不支持 root 用户直接启动elasticsearch

useradd es # 新建es用户

passwd es # 修改es用户的密码,输入2次密码回车就行

​

# passwd es

# Changing password for user es.

# New password:

# BAD PASSWORD: The password is shorter than 8 characters

# Retype new password:

# passwd: all authentication tokens updated successfully.

​

# 3、给es用户授权该目录

chown -R es:es /usr/local/elasticsearch-6.5.4   # -R 是指该文件下的全部文件都授权给es用户

​

# 4、max file descriptors [4096] for elasticsearch process is too low, increase to at least [65536] 超过启动shell的最大文件数，这个报错说明每个进程最大同时打开文件数太小

​

# 修改/etc/security/limits.conf文件，增加配置，用户退出后重新登录生效

vim /etc/security/limits.conf

​

# 在文件后面新增 es 的 这里不建议 把 es 修改成 \* 否则，对于其他用户都是放开限制

es         soft   nproc        65536

es         hard   nproc        65536

es         soft   nofile       65536

es         hard   nofile       65536

​

# 5、max virtual memory areas vm.max\_map\_count [65530] is too low, increase to at least [262144] 这个报错说明 vm 的启动内存太小，这个时候我们通过修改配置文件

vim /etc/sysctl.conf

​

# 使刚刚修改的命令生效

sysctl -p

​

# sysctl -p

net.ipv6.conf.all.disable\_ipv6 = 1

net.ipv6.conf.default.disable\_ipv6 = 1

net.ipv6.conf.lo.disable\_ipv6 = 1

vm.swappiness = 0

net.ipv4.neigh.default.gc\_stale\_time = 120

net.ipv4.conf.all.rp\_filter = 0

net.ipv4.conf.default.rp\_filter = 0

net.ipv4.conf.default.arp\_announce = 2

net.ipv4.conf.lo.arp\_announce = 2

net.ipv4.conf.all.arp\_announce = 2

net.ipv4.tcp\_max\_tw\_buckets = 5000

net.ipv4.tcp\_syncookies = 1

net.ipv4.tcp\_max\_syn\_backlog = 1024

net.ipv4.tcp\_synack\_retries = 2

kernel.sysrq = 1

vm.max\_map\_count = 262144l

​

# 6、上面的问题基本解决，那么我们就切换es用户启动elasticsearch就行

su - es

​

# 在Elasticsearch主目录下运行下列命令来安装这些插件：

bin/elasticsearch-plugin install ingest-geoip

bin/elasticsearch-plugin install ingest-user-agent

​

# 后台运行命令

./bin/elasticsearch -d

​

# 47.99.88.28 外网地址

$ curl  47.99.88.28:9200

{

 "name" : "node-1",

 "cluster\_name" : "elasticsearch",

 "cluster\_uuid" : "SaFb3ymvTAyDs6r0cpqkJQ",

 "version" : {

   "number" : "6.5.4",

   "build\_flavor" : "default",

   "build\_type" : "tar",

   "build\_hash" : "d2ef93d",

   "build\_date" : "2018-12-17T21:17:40.758843Z",

   "build\_snapshot" : false,

   "lucene\_version" : "7.5.0",

   "minimum\_wire\_compatibility\_version" : "5.6.0",

   "minimum\_index\_compatibility\_version" : "5.0.0"

},

 "tagline" : "You Know, for Search"

}

​

​**config/elasticsearch.yml**

cluster.name: elasticsearch # 集群名称

node.name: node-1 # 节点名称

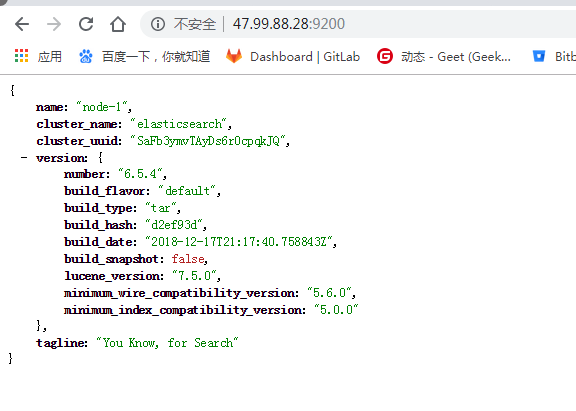
network.host: 172.16.229.30 # host 这里指定的是内网ip 可以用ip addr 命令查看，如果只有一张网卡的情况下，直接写0.0.0.0

http.port: 9200 # port 端口号

node.max\_local\_storage\_nodes: 2

http.cors.enabled: true

http.cors.allow-origin: "\*" # 跨域



# 7 logstash

filebeat中message要么是一段字符串，要么在日志生成的时候拼接成json然后在filebeat中指定为json。但是大部分系统日志无法去修改日志格式，filebeat则无法通过正则去匹配出对应的field，这时需要结合logstash的grok来过滤。Logstash 是开源的服务器端数据处理管道，能够同时从多个来源采集数据，转换数据，然后将数据发送到存储库中

# 解压到 /usr/local/ 对应的目录下

tar -zxvf logstash-6.5.4.tar.gz -C /usr/local/

​

# 进入 logstash 目录

cd /usr/local/logstash-6.5.4/

​

# 注: Logstash实现的功能主要分为接收数据、解析过滤并转换数据、输出数据三个部分，对应的插件依次是input插件、filter插件、output插件，其中，filter插件是可选的，其它两个是必须插件。也就是说在一个完整的Logstash配置文件中，必须有input插件和output插件。

cd config/

​

# 创建 logstash 配置文件 有一个模板文件 logstash-sample.conf

vim logstash.conf

​

# 创建 logstash 配置文件 有一个模板文件 logstash-sample.conf

mkdir -p /usr/local/logstash-6.5.4/patterns

cd /usr/local/logstash-6.5.4/patterns

vi java

# user-center

MYAPPNAME ([0-9a-zA-Z\_-]\*)

MYTHREADNAME ([0-9a-zA-Z.\_-]|\(|\)|\s)\*

:wq

chmod -r 777 /usr/local/logstash-6.5.4/patterns

# 回到原目录，在启动 logstash

cd ..

./bin/logstash -f config/logstash.conf

​

# 8 kibana

通过 Kibana，能够对 Elasticsearch 中的数据进行可视化并在 Elastic Stack 进行操作.

# 解压到 /usr/local/ 对应的目录下

tar -zxvf kibana-6.5.4-linux-x86\_64.tar.gz -C /usr/local/

​

# 进入 kibana 目录

cd /usr/local/kibana-6.5.4-linux-x86\_64/

​

# 编辑 kibana 配置文件

vim config/kibana.yml

​

# 配置文件中 找到对应的属性

server.host: "172.16.229.30" # 内网ip 可以写成0.0.0.0

elasticsearch.url: "http://172.16.229.30:9200" # elasticsearch的地址

​

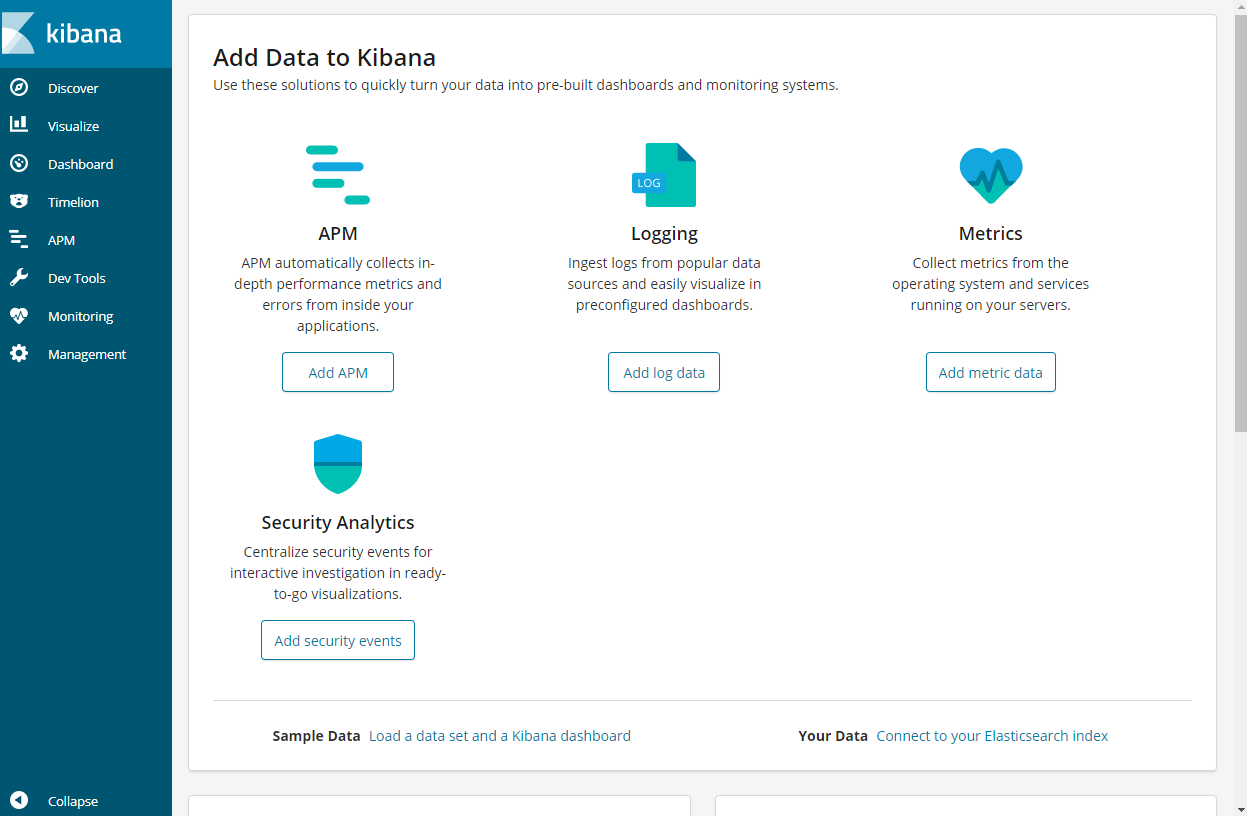
# 启动 kibana

./bin/kibana

​

# 9 查看日志

使用浏览器访问[http://192.168.33.10:5601](http://192.168.33.10:5601/)便可以看到Kibana管理界面：



## 修改微服务日志配置

就目前微服务模块来说向pom.xml里引入Logstash依赖：

<dependency>

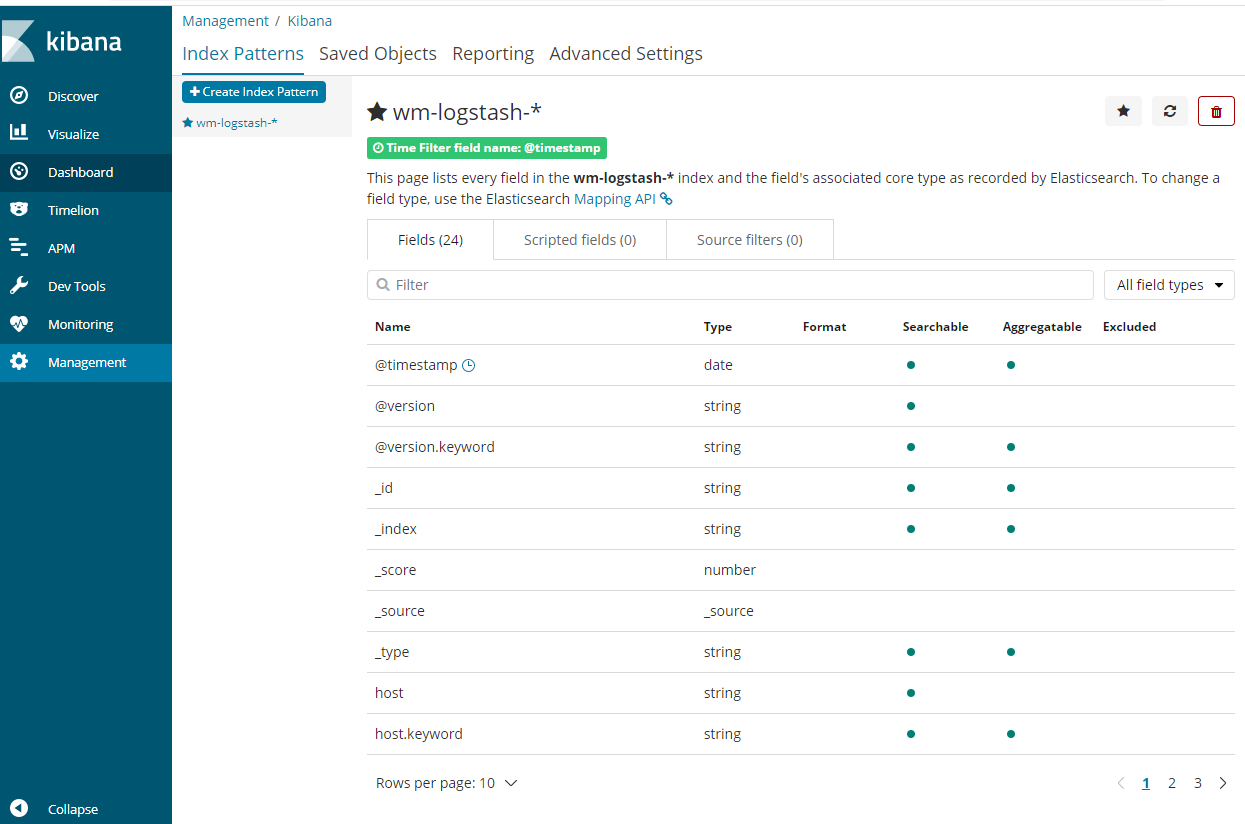
<groupId>net.logstash.logback</groupId>

<artifactId>logstash-logback-encoder</artifactId>

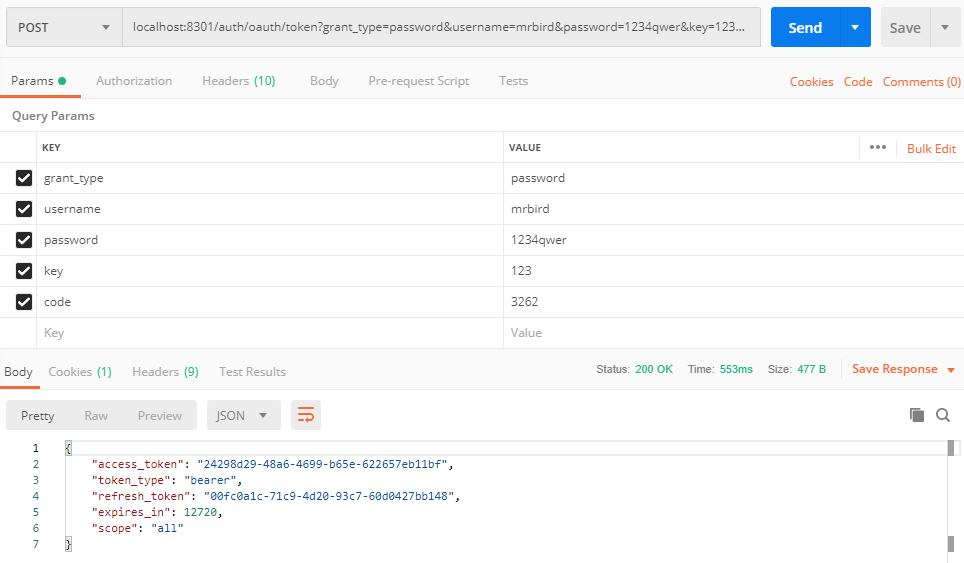
<version>6.1</version>

</dependency>

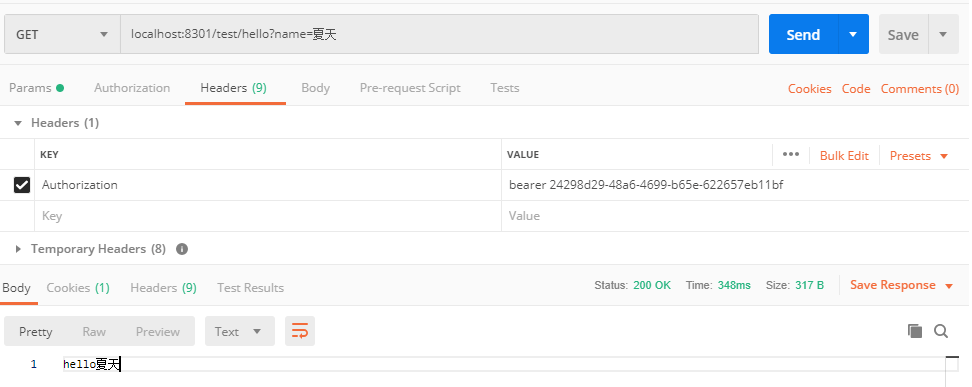
在Index pattern里输入我们在logstash配置文件logstash-wm.conf里output.index指定的值wm-logstash-\*：



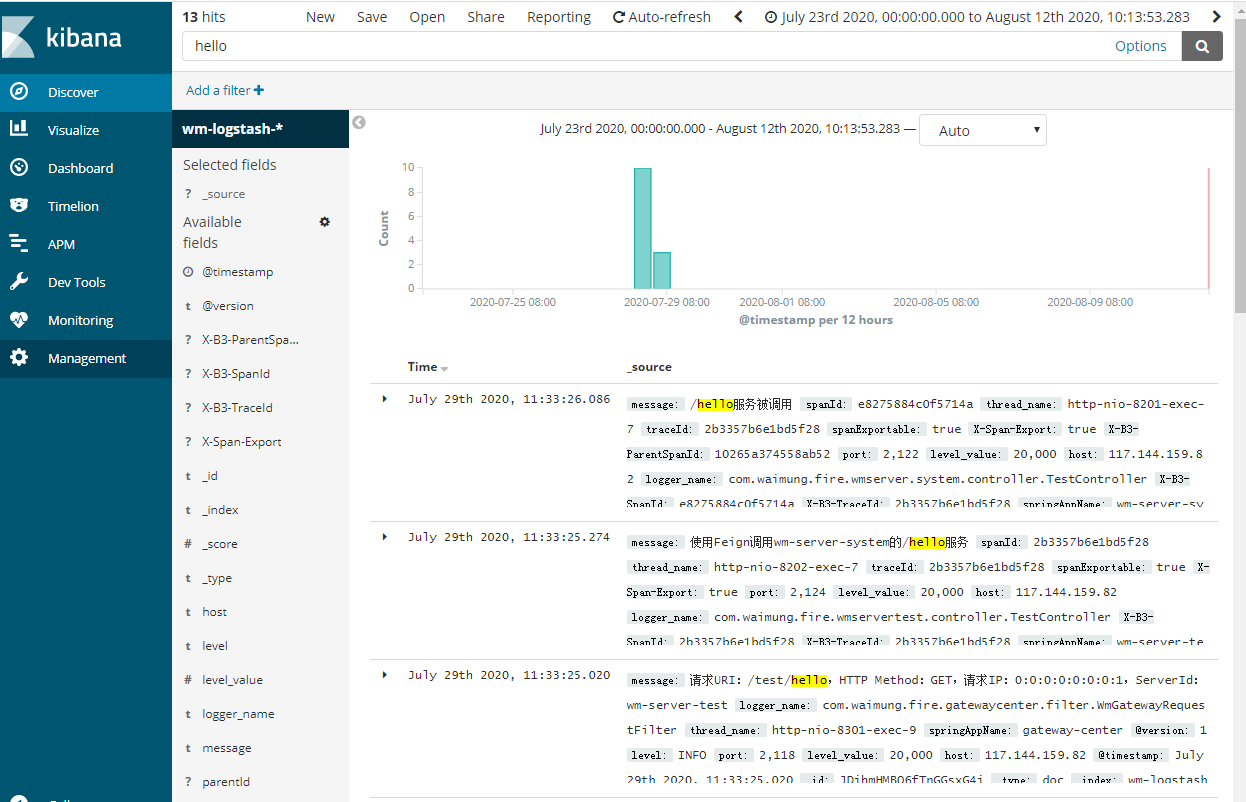
使用PostMan获取令牌：



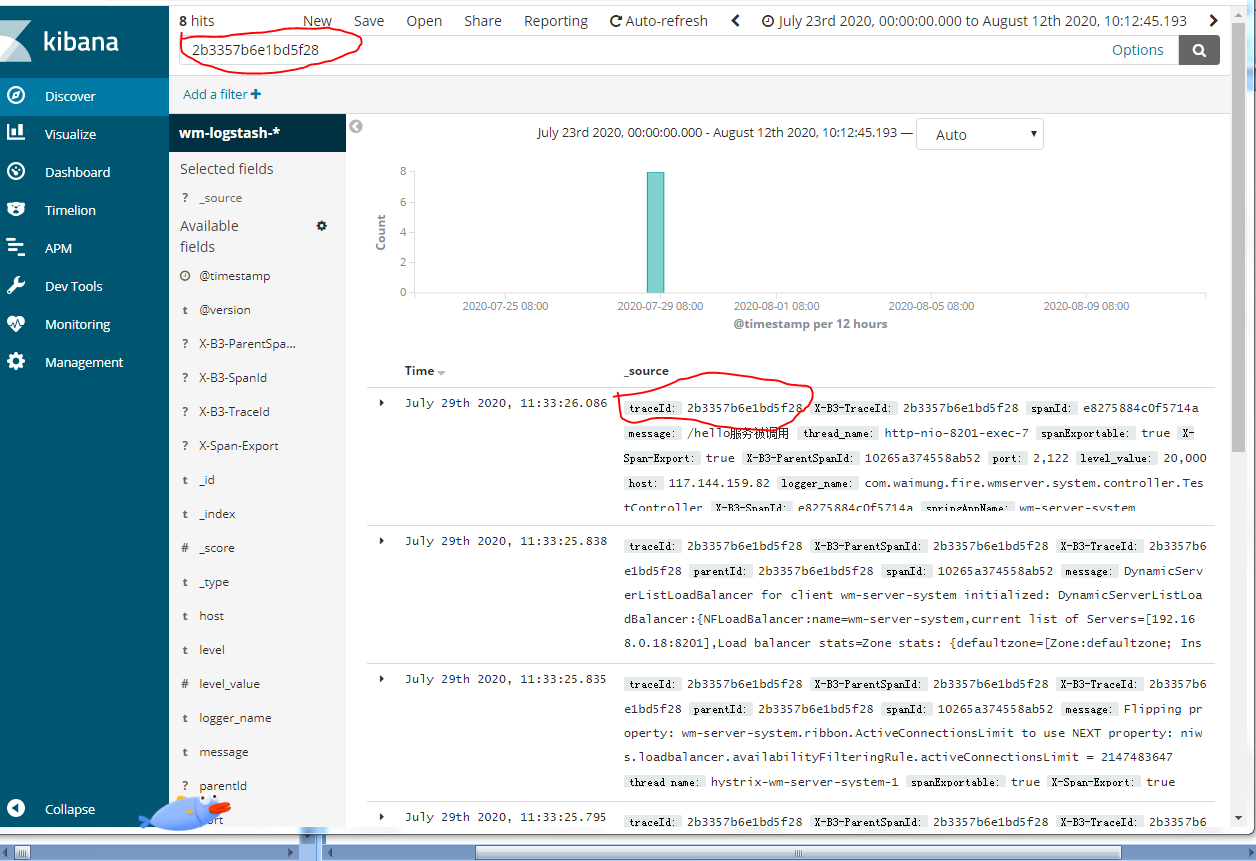
然后发送一笔[localhost:8301/test/hello?name=夏天](localhost:8301/test/hello?name=%E5%A4%8F%E5%A4%A9)请求：



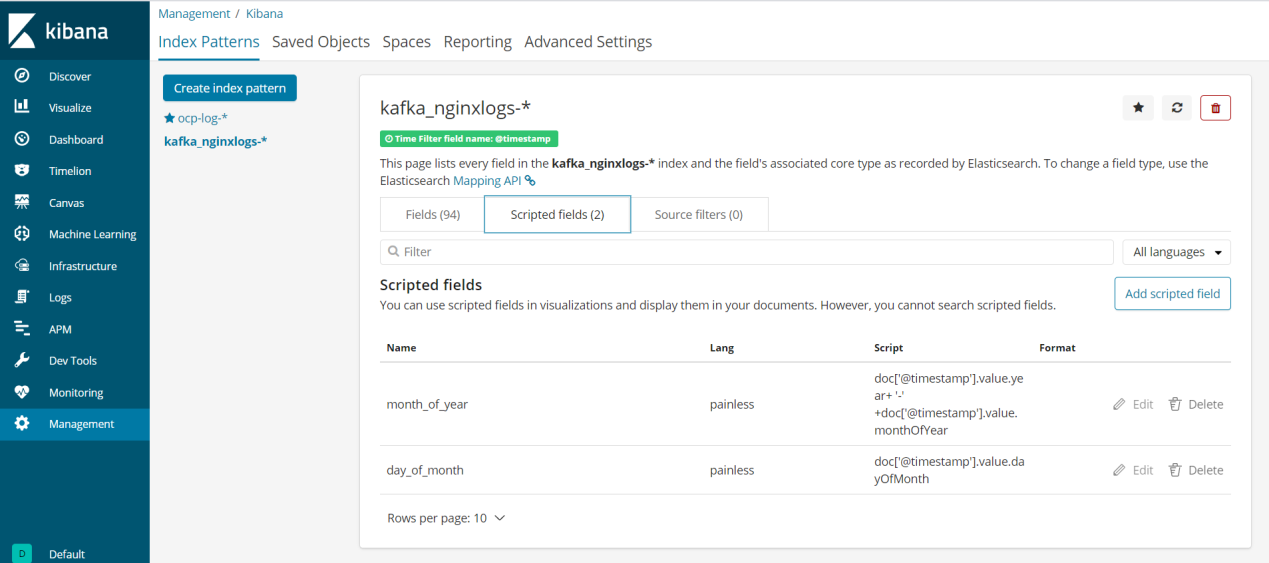
在Kibana里搜索这笔日志：



搜索traceid



# 10实战



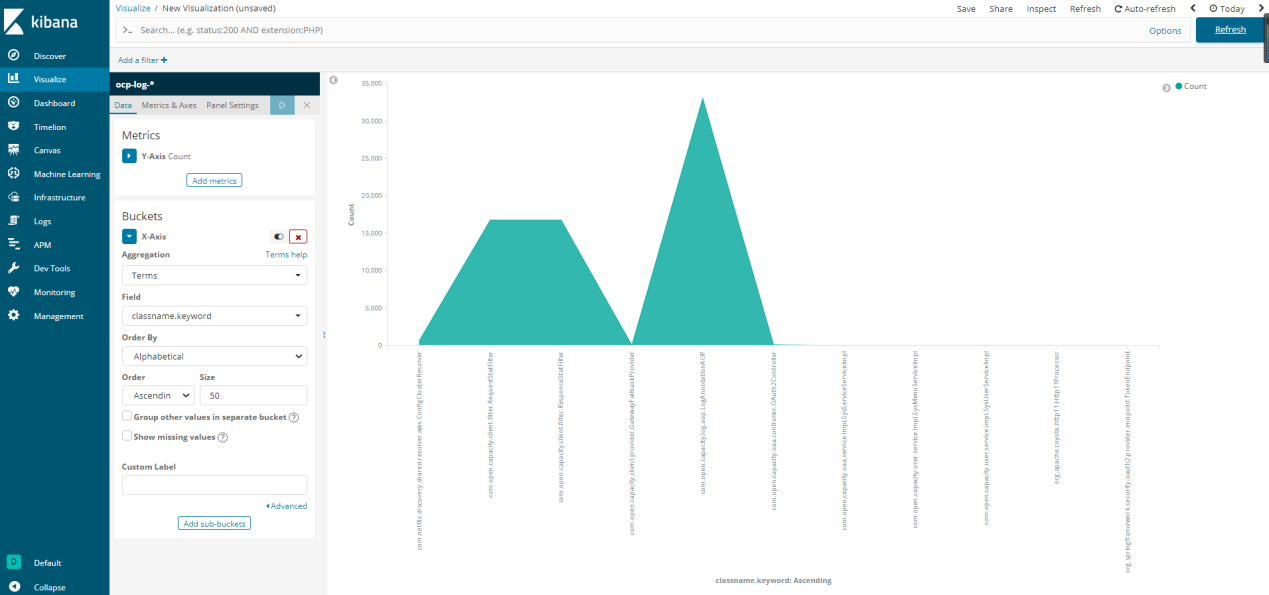
month\_of\_year: doc['@timestamp'].value.year+ '-' +doc['@timestamp'].value.monthOfYear

day\_of\_month: doc['@timestamp'].value.dayOfMonth

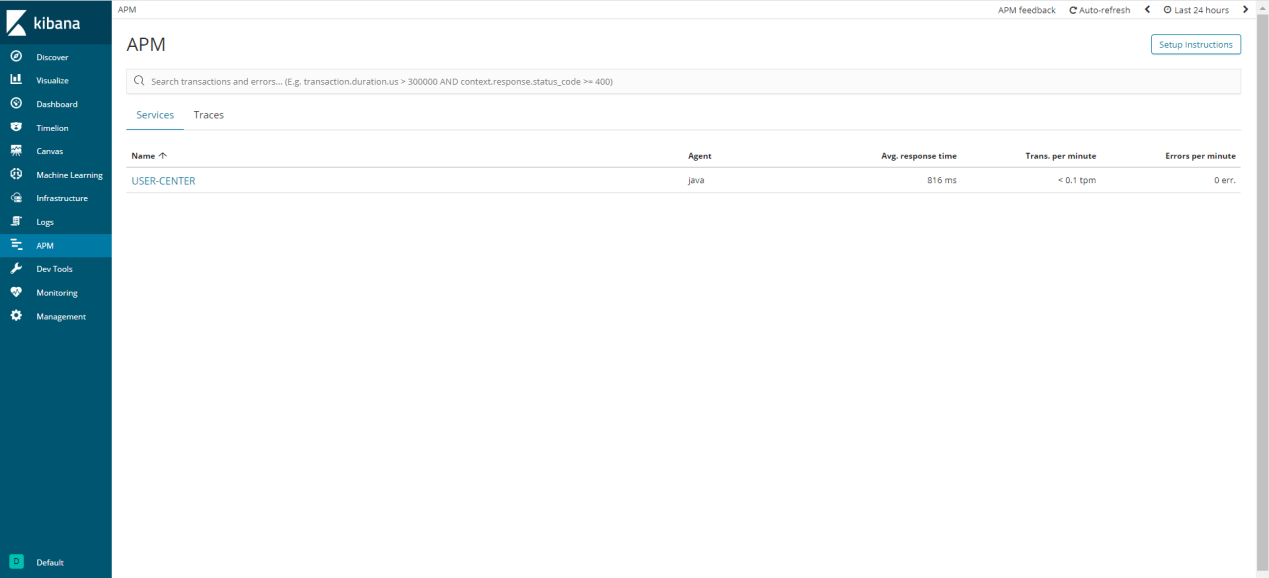
按月，日统计nginx流量：



统计项目日志中调用频繁的类



# 11 APM 监控



## 搭建apm-server

wget <https://artifacts.elastic.co/downloads/apm-server/apm-server-6.5.4-linux-x86_64.tar.gz>

[root@demo03 apm-server-6.5.4-linux-x86\_64]# vi apm-server.yml

apm-server:

# Defines the host and port the server is listening on. use "unix:/path/to.sock" to listen on a unix domain socket.

host: "0.0.0.0:8200"

run:

enabled: true

output.elasticsearch:

# Array of hosts to connect to.

# Scheme and port can be left out and will be set to the default (http and 9200)

# In case you specify and additional path, the scheme is required: http://localhost:9200/path

# IPv6 addresses should always be defined as: https://[2001:db8::1]:9200

hosts: ["47.99.88.28:9200"]

indices:

- index: "apm-%{[beat.version]}-sourcemap"

when.contains:

processor.event: "sourcemap"

- index: "apm-%{[beat.version]}-error-%{+yyyy.MM.dd}"

when.contains:

processor.event: "error"

- index: "apm-%{[beat.version]}-transaction-%{+yyyy.MM.dd}"

when.contains:

processor.event: "transaction"

- index: "apm-%{[beat.version]}-span-%{+yyyy.MM.dd}"

when.contains:

processor.event: "span"

- index: "apm-%{[beat.version]}-metric-%{+yyyy.MM.dd}"

when.contains:

processor.event: "metric"

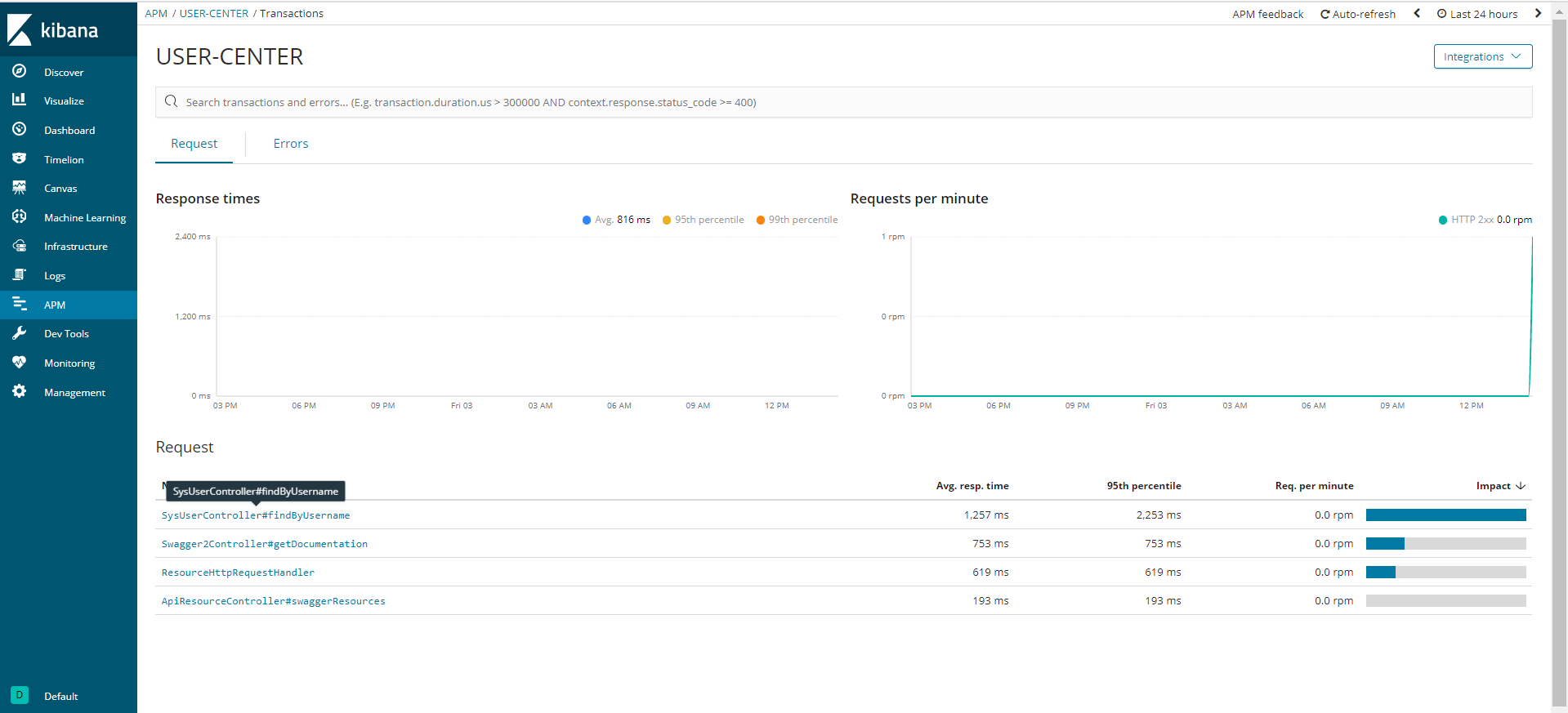
- index: "apm-%{[beat.version]}-onboarding-%{+yyyy.MM.dd}"

when.contains:

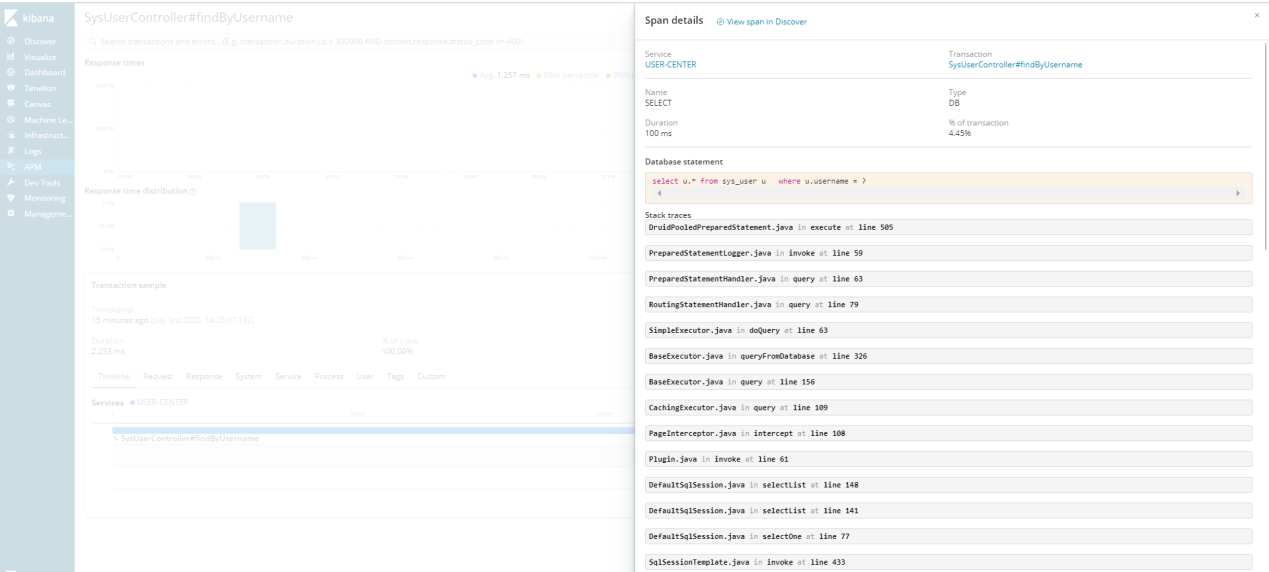
processor.event: "onboarding"

启动 nohup ./apm-server -e >&/dev/null &

## 监控大盘



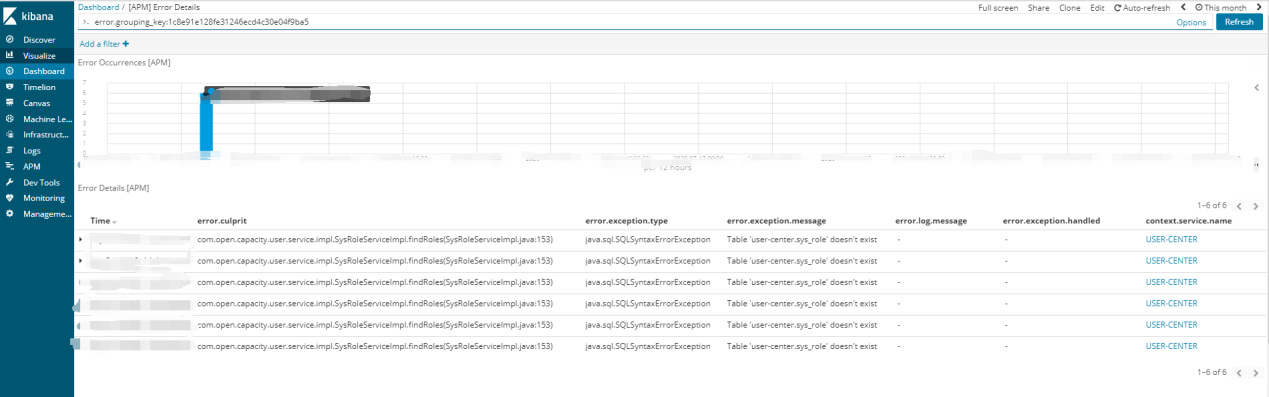
## Sql 语句



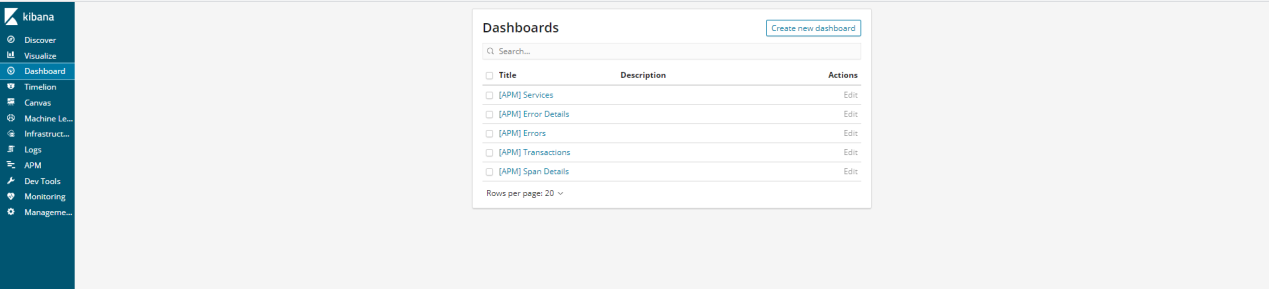
## 错误列表



## 错误统计



## Dashboard



## Service

