

Kafka+ELK

概述

ELK (Elasticsearch、Logstash、Kibana) + Kafka来搭建一个日志系统。

什么是ELK?

Elasticsearch是一个基于Lucene、分布式、通过Restful方式进行交互的近实时搜索平台框架。像类似百度、谷歌这种大数据全文搜索引擎的场景都可以使用Elasticsearch作为底层支持框架，可见Elasticsearch提供的搜索能力确实强大，市面上很多时候我们简称Elasticsearch为es。

当然，Elasticsearch 不仅仅是 Lucene，并且也不仅仅只是一个全文搜索引擎。它可以被下面这样准确地形容：

- 一个分布式的实时文档存储，每个字段可以被索引与搜索；
- 一个分布式实时分析搜索引擎；
- 能胜任上百个服务节点的扩展，并支持 PB 级别的结构化或者非结构化数据。

Logstash是ELK的中央数据流引擎，用于从不同目标（文件/数据存储/MQ）收集的不同格式数据，经过过滤后支持输出到不同目的地（文件/MQ/redis/Elasticsearch/Kafka等）。

Kibana可以将Elasticsearch的数据通过友好的页面展示出来，提供实时分析的功能。

为什么用ELK?

1. 以前不用ELK的做法

一般单体结构的项目使用log4j来把日志写到log文件中。

微服务之后，项目有了高可用的要求，进行了分布式部署web。如果我们还是用log4j这样的方式来记录log的话，那么有多少个分布式机器，就有多少个日志记录，这个时候查找log起来非常麻烦，不方便定位bug。后来，直接将log写到数据库中去，这样做，虽然解决了查找异常信息便利性的问题了，但存在两个缺陷：

1. log记录一多，表不够用，必须分库分表
2. 使用数据库必须考虑到数据库的异常，如果数据库异常，log就会出现丢失了。那么为了解决log丢失的问题，那么还得先将log写在本地，然后等db连通了后，再将log同步到db。

2. 现在ELK的做法

ELK方案，可以解决以上问题。

首先是，使用Elasticsearch来存储日志信息，对一般系统来说可以理解为可以存储无限条数据，因为Elasticsearch有良好的扩展性，然后是有一个Logstash，可以把理解为数据接口，为Elasticsearch对接外面过来的log数据，它对接的渠道，有Kafka、log、redis等等，最后还有一个部分就是kibana，它主要用来做数据展现，log那么多数据都存放在Elasticsearch中，需要可视化展示，这个kibana就是为了让我们的看log数据的，但还有一个更重要的功能是，可以编辑N种图表形式，什么柱状图，折线图等等，来对log数据进行直观的展现。

业务流程：

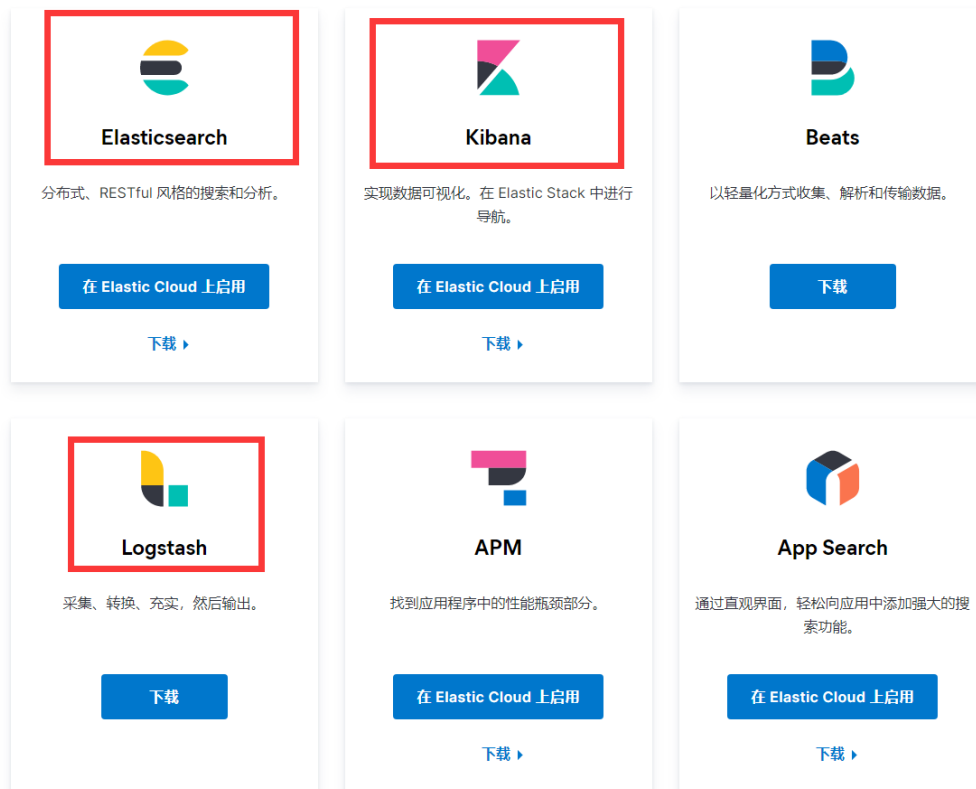
1. 使用Spring aop进行日志收集

2. 通过Kafka将日志发送给Logstash
3. Logstash再将日志写入Elasticsearch，这样Elasticsearch就有了日志数据了。
4. 使用Kibana将存放在Elasticsearch中的日志数据显示出来，实时的数据图表分析。




ELK搭建

1. 下载ElasticSearch+Logstash+Kibana

官网地址<https://www.elastic.co/downloads>



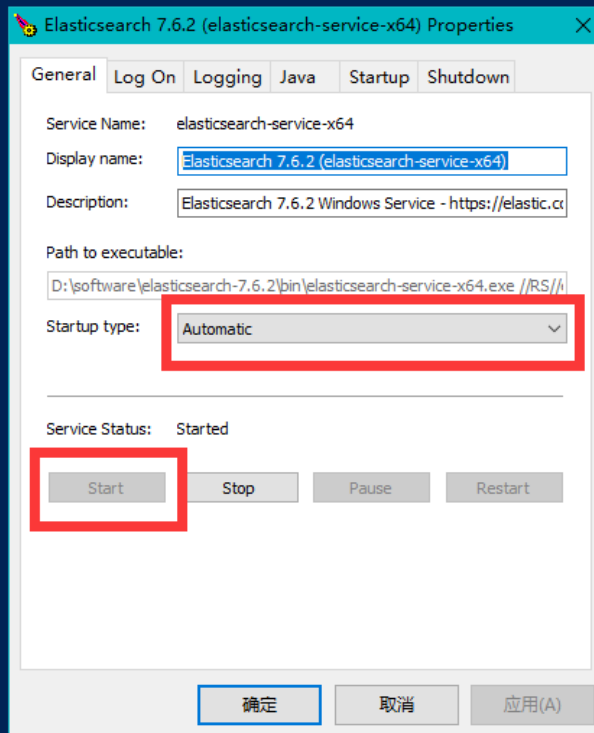
下载后解压

 elasticsearch-7.6.2-windows-x86_64.zip	2020/4/27 9:25	压缩(zipped)文件...	286,268 KB
 kibana-7.6.2-windows-x86_64.zip	2020/4/27 9:27	压缩(zipped)文件...	290,699 KB
 logstash-7.6.2.zip	2020/4/27 9:24	压缩(zipped)文件...	175,362 KB

2. 启动 Elasticsearch

- 打开elasticsearch-7.6.2\bin，cmd运行elasticsearch-service.bat install
- 运行 elasticsearch-service.bat manager 管理配置ES，点击Start启动服务

```
S D:\software\elasticsearch-7.6.2\bin> ./elasticsearch-service.bat manager
```



- 这里可以设置automatic开机自启。
- 输入网址 <http://localhost:9200/>，可以看到如下信息

```
< > ↺ 🏠 📖 | ☆ localhost:9200

{
  "name" : "DESKTOP-T5Q6LSJ",
  "cluster_name" : "elasticsearch",
  "cluster_uuid" : "tGYFisntRFSGKv8h-ZJbPA",
  "version" : {
    "number" : "7.6.2",
    "build_flavor" : "default",
    "build_type" : "zip",
    "build_hash" : "ef48eb35cf30adf4db14086e8aabd07ef6fb113f",
    "build_date" : "2020-03-26T06:34:37.794943Z",
    "build_snapshot" : false,
    "lucene_version" : "8.4.0",
    "minimum_wire_compatibility_version" : "6.8.0",
    "minimum_index_compatibility_version" : "6.0.0-beta1"
  },
  "tagline" : "You Know, for Search"
}
```

其中会有版本号等信息。

3. Logstash

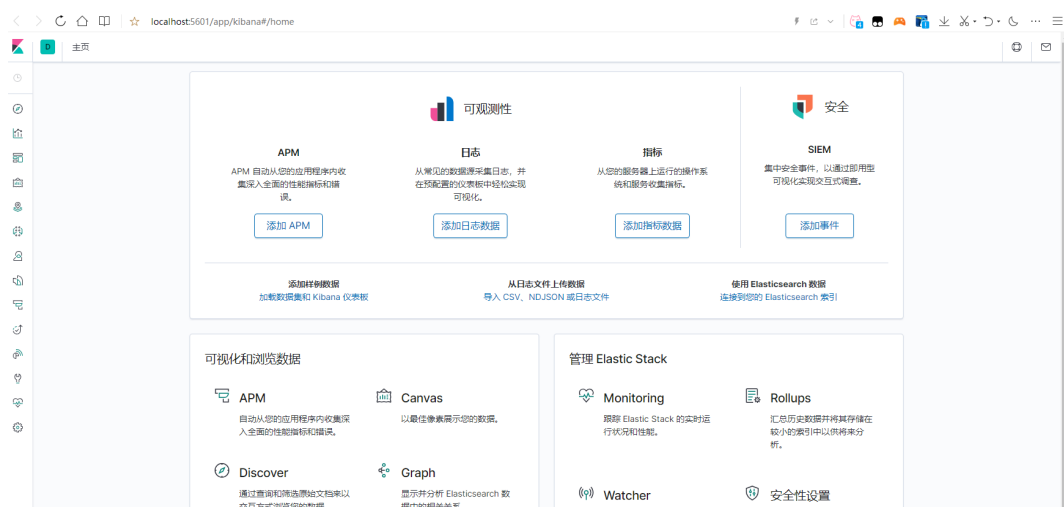
Logstash是一个导入导出数据的工具，使用时直接编辑配置文件，输入指令即可。下面会结合实例介绍具体用法。

4. 启动Kibana

- 进入kibana-7.6.2-windows-x86_64\bin，双击运行kibana.bat

```
C:\Windows\system32\cmd.exe
log [02:08:24.815] [info] [status] [plugin:oss_telemetry@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.819] [info] [status] [plugin:file_upload@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.823] [info] [status] [plugin:data@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.830] [info] [status] [plugin:lens@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.843] [info] [status] [plugin:snapshot_restore@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.850] [info] [status] [plugin:input_control_vis@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.852] [info] [status] [plugin:kibana_react@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.853] [info] [status] [plugin:management@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.855] [info] [status] [plugin:navigation@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.857] [info] [status] [plugin:region_map@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.863] [info] [status] [plugin:telemetry@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.985] [info] [status] [plugin:timelion@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.989] [info] [status] [plugin:ui_metric@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.990] [info] [status] [plugin:markdown_vis@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.991] [info] [status] [plugin:metric_vis@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.993] [info] [status] [plugin:vega@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.995] [info] [status] [plugin:tagcloud@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:24.996] [info] [status] [plugin:table_vis@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:26.507] [warning] [reporting] 正在为 xpack.reporting.encryptionKey 生成随机密钥。要防止待处理报告在重新启动时失败, 请在 kibana.yml 中设置 xpack.reporting.encryptionKey
log [02:08:26.513] [info] [status] [plugin:reporting@7.6.2] Status changed from uninitialized to green - Ready
log [02:08:26.532] [info] [listening] Server running at http://localhost:5601
log [02:08:26.636] [info] [server] [Kibana] [http] http server running at http://localhost:5601
Could not get dynamic index pattern because indices "apm-*" don't exist
Could not get dynamic index pattern because indices "apm-*" don't exist
Could not get dynamic index pattern because indices "apm-*" don't exist
```

- 可以访问 <http://localhost:5601> 查看是否启动成功

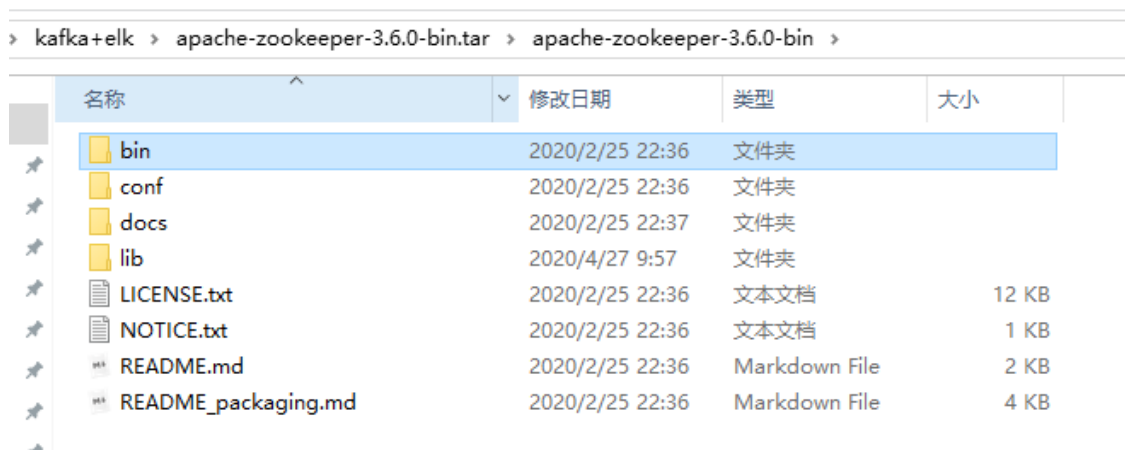


Kafka搭建

1. 安装ZooKeeper

因为Kafka依赖于ZooKeeper, 所以先安装Zookeeper。

- 下载ZooKeeper, <https://downloads.apache.org/zookeeper/zookeeper-3.6.0/apache-zookeeper-3.6.0-bin.tar.gz>
- 解压缩



- 修改配置文件

\apache-zookeeper-3.6.0\conf下, 重命名zoo_sample.cfg为zoo.cfg

```

1 # The number of milliseconds of each tick
2 tickTime=2000
3 # The number of ticks that the initial
4 # synchronization phase can take
5 initLimit=10
6 # The number of ticks that can pass between
7 # sending a request and getting an acknowledgement
8 syncLimit=5
9 # the directory where the snapshot is stored.
10 # do not use /tmp for storage, /tmp here is just
11 # example sake.
12 dataDir=D:\software\zookeeper_log
13 # the port at which the clients will connect
14 clientPort=2181
15 # the maximum number of client connections.
16 # increase this if you need to handle more clients
17 #maxClientCnxns=60
18 #
19 # Be sure to read the maintenance section of the
20 # administrator guide before turning on autopurge.
21 #
22 # http://zookeeper.apache.org/doc/current/zookeeperAdmin.html#sc_m

```

- ```
C:\Windows\system32\cmd.exe
D:\software\apache-zookeeper-3.6.0-bin\bin>call "C:\Program Files\Java\jdk1.8.0.191\bin\java -Dzookeeper.log.dir=D:\software\apache-zookeeper-3.6.0-bin\bin\.\logs -Dzookeeper.root.logger=INFO, CONSOLE" -Dzookeeper.log.file=zookeeper-CC-server-DESKTOP-T506LSJ.log" -XX:HeapDumpOnOutOfMemoryError" -XX:OnOutOfMemoryError=cmd /c taskkill /pid %p /t /f" -cp "D:\software\apache-zookeeper-3.6.0-bin\bin\.\build\classes;D:\software\apache-zookeeper-3.6.0-bin\bin\.\build\lib*" D:\software\apache-zookeeper-3.6.0-bin\bin\.**;D:\software\apache-zookeeper-3.6.0-bin\bin\.\lib*;D:\software\apache-zookeeper-3.6.0-bin\bin\.\conf org.apache.zookeeper.server.quorum.QuorumPeerMain "D:\software\apache-zookeeper-3.6.0-bin\bin\.\conf\zoo.cfg"
2020-04-28 09:55:37,002 [myid:] - INFO [main:QuorumPeerConfig@173] - Reading configuration from: D:\software\apache-zookeeper-3.6.0-bin\bin\.\conf\zoo.cfg
2020-04-28 09:55:37,020 [myid:] - WARN [main:VerifyingFileFactory@65] - D:\software\apache-zookeeper-3.6.0-bin\bin\.\conf\zoo.cfg to indicate that you're sure!
2020-04-28 09:55:37,026 [myid:] - INFO [main:QuorumPeerConfig@459] - clientPortAddress is 0.0.0.0:2181
2020-04-28 09:55:37,028 [myid:] - INFO [main:QuorumPeerConfig@463] - secureClientPort is not set
2020-04-28 09:55:37,028 [myid:] - INFO [main:QuorumPeerConfig@479] - observerMasterPort is not set
2020-04-28 09:55:37,028 [myid:] - INFO [main:QuorumPeerConfig@496] - metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider
2020-04-28 09:55:37,047 [myid:] - INFO [main:DataDirCleanupManager@78] - autopurge.snapRetainCount set to 3
2020-04-28 09:55:37,047 [myid:] - INFO [main:DataDirCleanupManager@79] - autopurge.purgeInterval set to 0
2020-04-28 09:55:37,047 [myid:] - INFO [main:DataDirCleanupManager@101] - Purge task is not scheduled.
2020-04-28 09:55:37,047 [myid:] - WARN [main:QuorumPeerMain@138] - Either no config or no quorum defined in config, running in standalone mode
2020-04-28 09:55:37,049 [myid:] - INFO [main:ManagedUtil@45] - Log4j found with jmx enabled.
2020-04-28 09:55:37,112 [myid:] - INFO [main:QuorumPeerConfig@173] - Reading configuration from: D:\software\apache-zookeeper-3.6.0-bin\bin\.\conf\zoo.cfg
2020-04-28 09:55:37,113 [myid:] - WARN [main:VerifyingFileFactory@65] - D:\software\apache-zookeeper-3.6.0-bin\bin\.\conf\zoo.cfg to indicate that you're sure!
2020-04-28 09:55:37,113 [myid:] - INFO [main:QuorumPeerConfig@459] - clientPortAddress is 0.0.0.0:2181
2020-04-28 09:55:37,113 [myid:] - INFO [main:QuorumPeerConfig@463] - secureClientPort is not set
2020-04-28 09:55:37,113 [myid:] - INFO [main:QuorumPeerConfig@479] - observerMasterPort is not set
```

```
C:\Users\CQ>jps
1232 AuthBootstrap
18752 ConsoleProducer
1156 Kafka
20852
10840 QuorumPeerMain
1736 Logstash
21976 Launcher
7832 BasicSystemBootstrap
21580 GatewayBootstrap
5004 Jps
556 RemoteMavenServer

C:\Users\CQ>
```

- o 下载地址<http://kafka.apache.org/downloads.html>
- o 修改配置文件，进入kafka\_2.12-2.5.0\config，修改server.properties



