### EFK日志搜集demo

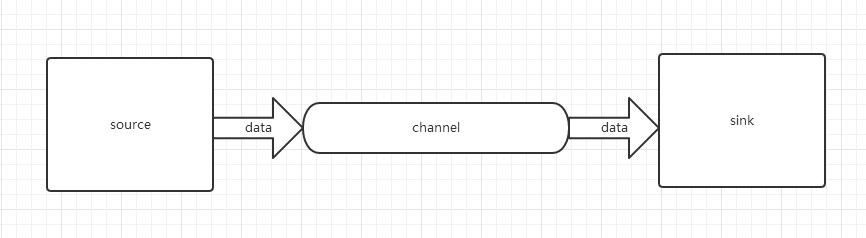
E：Elasticsearch

F：Flume

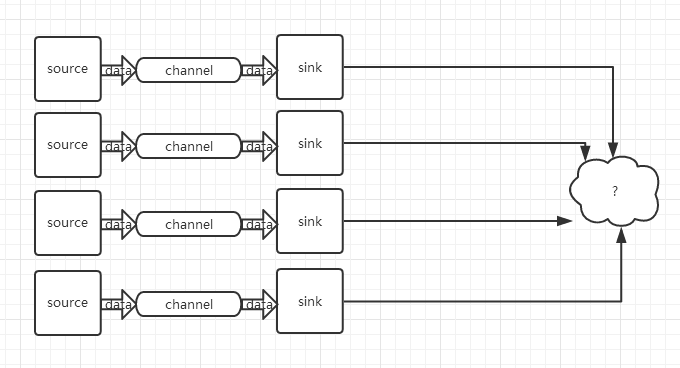
K：Kafka

Flume是一个分布式的日志聚合收集工具，可以从多个且不同类型的日志源头收集日志。Flume的模型如下：

Source代表数据的源头，channel暂存数据，sink为数据的流向。如下：



多个flume代理的情况下，数据可以汇聚到同一个地方，如下：



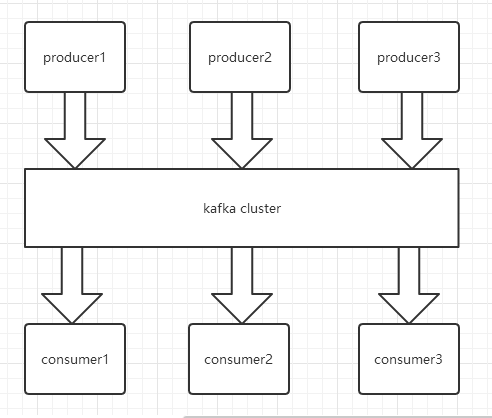
数据量多的时候，可能终端的数据处理压力比较大，为了平衡数据生产的速度和处理速度，最好在数据生产和数据处理之间增加一个缓冲，为此使用Kafka。

Apache Kafka是一种分布式发布-订阅消息系统。

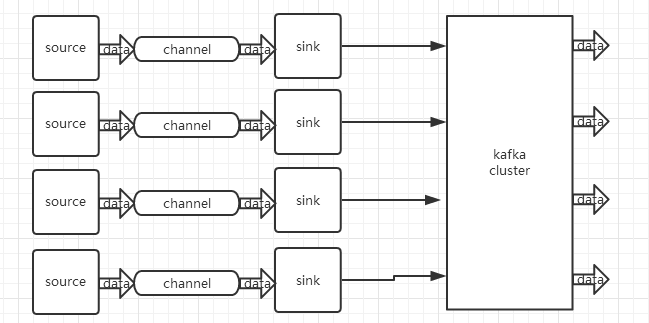
kafka的设计初衷是希望作为一个统一的信息收集平台,能够实时的收集反馈信息,并需要能够支撑较大的数据量,且具备良好的容错能力。

Kafka是一个典型的生产者消费者模型，生产者消费者模型的主要功能就是提供缓冲，平衡数据生产和使用的速率。

Kafka的模型如下：



所以Flume负责收集数据，Kafka作为缓冲，消费者从kafka中取数据。就像是多个水管同时向池子中注水，水管从哪取水可以自己决定；同时也有多个水管从池子里取水，水要流向哪里也可以自己决定。



Elasticsearch是一个基于Lucene的搜索服务器，提供了分布式的全文搜索引擎，传统的数据库如果数据在PB（1024TB）级别的，搜索会非常慢，而Elasticsearch支持PB级别的搜索。

Elasticsearch的几个概念：

Cluster：集群，一个集群下有多个节点。

Node：相同的集群名的节点是一个集群。

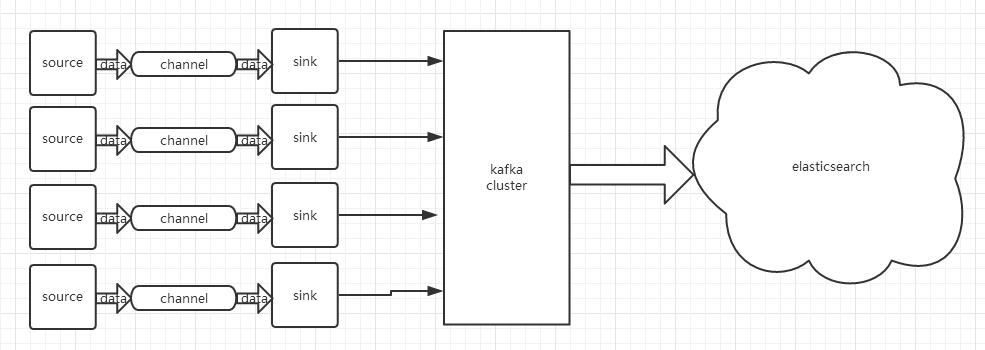
Index：Elasticsearch用来存储数据的逻辑区域。

Document：文档，数据实体。

Document Type：文档的类型。

所以可以为了搜索的方便，可以把kafka的数据存入Elasticsearch。

模型如下：



现在建立一个简单的日志收集搜索的demo：

jdk要求版本：1.7

①先简单的捕获异常并记录，规范异常的记录格式如下：记录异常的各种信息。因为elasticsearch的文档都采用json的格式，所以异常规范了扔到Elasticsearch里。

{

"localizedMessage": "分页查询Activity对象",

"cause": "java.lang.NullPointerException",

"mac": "",

"type": "java.lang.RuntimeException",

"extendedStackTrace": [

{

"author": "风中追风",

"file": "ActivityManagerAction.java",

"traceSeat": 1,

"line": 202,

"class": "com.cairh.xpe.coupons.backend.action.ActivityManagerAction",

"method": "getActivityList"

}

],

"ip": "127.0.0.1",

"message": "分页查询Activity对象",

"time": "2017-10-19 09:25:07",

"questParams": {

"page": [

"1"

],

"rtQryPage": [

"1"

],

"activity\_type": [

"2"

]

},

"serverPort": 80,

"thread": "http-apr-8080-exec-10",

"remotePort": 55192,

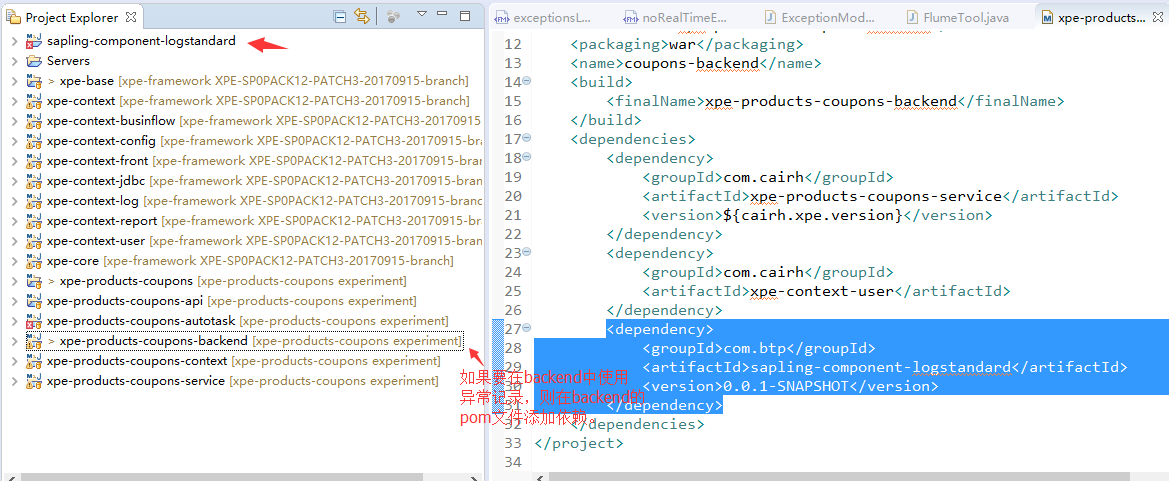
"URL": "http://hcsisap.xpe.com/xpe-products-coupons-backend/coupons/getActivityList.json",

"timemillis": 1508376307761,

"URI": "/xpe-products-coupons-backend/coupons/getActivityList.json"

}

②导入sapling-component-logstandard项目，且在要使用异常记录的项目的pom文件添加依赖：



<dependency>

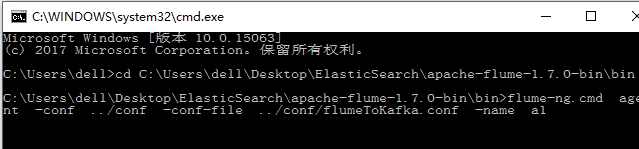
<groupId>com.btp</groupId>

<artifactId>sapling-component-logstandard</artifactId>

<version>0.0.1-SNAPSHOT</version>

</dependency>

③启动Flume实例：

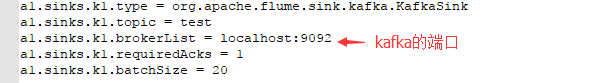


在bin目录下执行cmd：flume-ng.cmd agent -conf ../conf -conf-file ../conf/flumeToKafka.conf -name a1

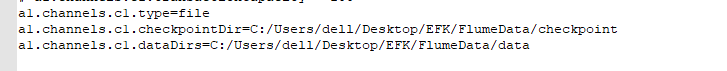
flumeToKafka.conf为启动一个flume代理的配置文件，可以配置flume的source、channel和sink，a1为代理名称。比如



配置source：代表从44444端口获取数据。

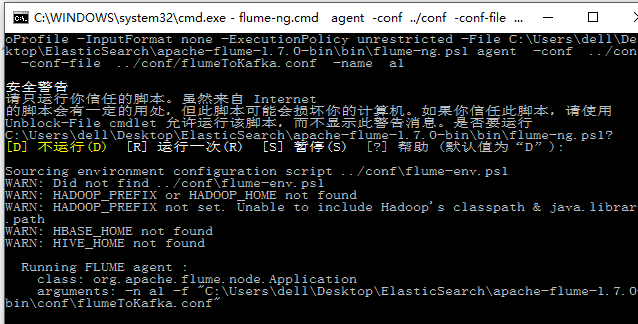


配置sink，传数据到kafka的名为test的topic下。



配置channel：本地的flume数据缓存。

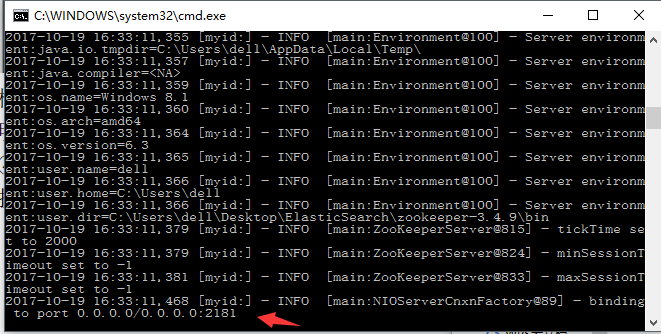
启动代理之后：



检查一下上面配置的44444端口是否打开：telnet localhost 44444,如果不是报未打开，则说明flume代理成功启动。

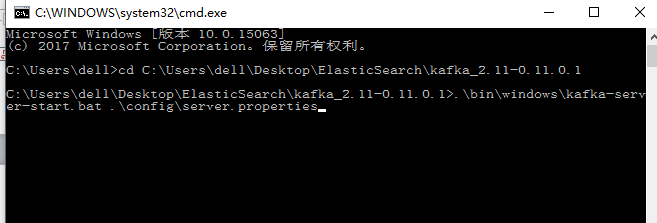
④启动zookeeper，因为kafka依赖zookeeper，所以先启动zookeeper：

执行zookeeper-3.4.9\bin\zkServer.cmd



⑤启动kafka：

Kafka目录下启动：.\bin\windows\kafka-server-start.bat .\config\server.properties

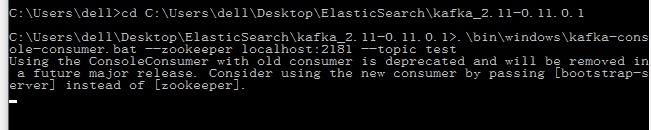


上面的flume的sink的9092端口配置在producer.properties。

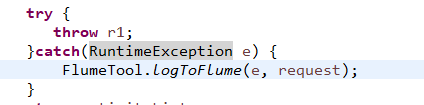
启动一个消费者：

消费者（Kafka目录下启动）：.\bin\windows\kafka-console-consumer.bat --zookeeper localhost:2181 --topic test

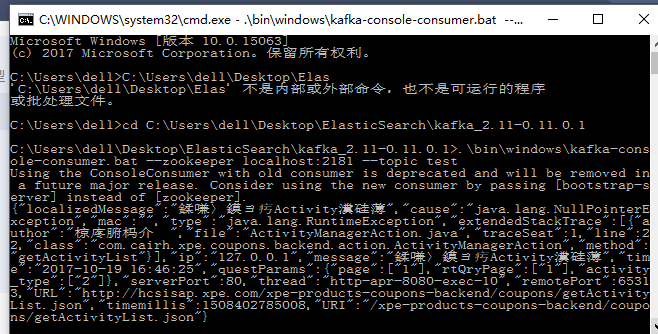
消费topic为test的数据：



然后记录日志：



查看kafka消费端窗口：已经有异常消息了。



⑥准备将消息送到elasticsearch中：

在backend的web.xml增加

<servlet-name>action</servlet-name>

<servlet-class>com.btp.logstandard.autoTask.KafkaComsumer</servlet-class>

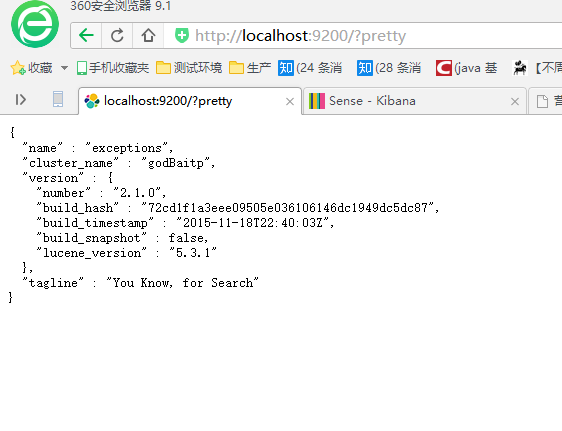
<load-on-startup>2</load-on-startup>

</servlet>

启动elasticsearch：elasticsearch-2.1.0\bin\elasticsearch.bat

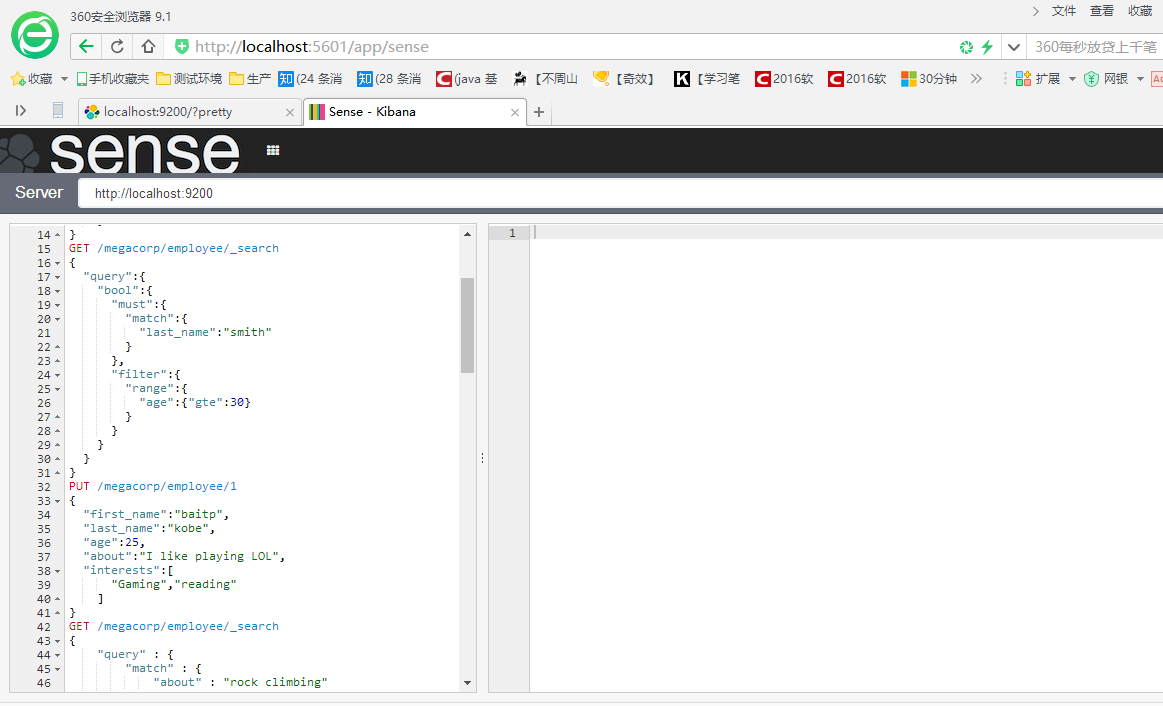
elasticsearch-2.1.0\config\elasticsearch.yml可以配置集群名称和节点名称。

打开网页：输入：http://localhost:9200/?pretty

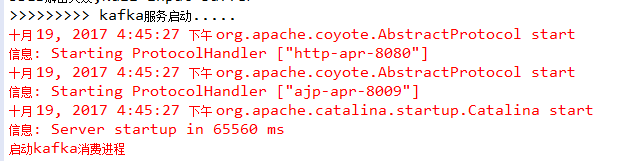
说明启动成功。

启动kibana：kibana-4.3.0-windows\bin\kibana.bat

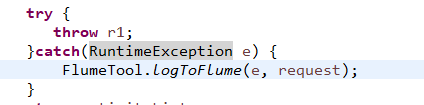
打开网页输入：<http://localhost:5601/app/sense>

表示启动kinaba成功。

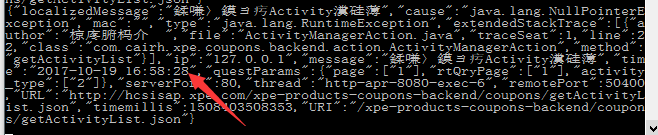
⑦启动web.xml改变的项目：

如果如下说明启动成功，现在可以往elasticsearch中放数据了。

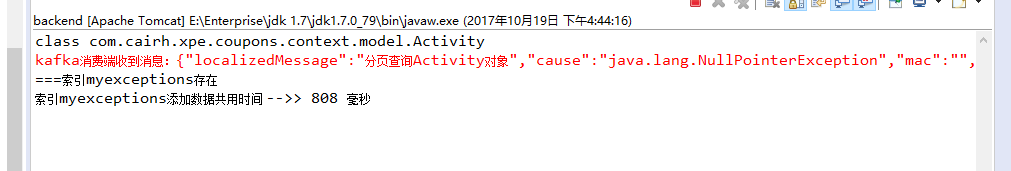
再次调用



Kafka消费窗口可看到：

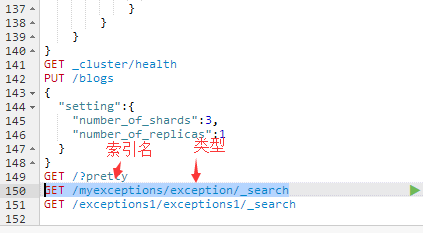


Console下（默认存到myexceptions索引下，类型定义为exception）：



说明数据存到elasticsearch中成功。

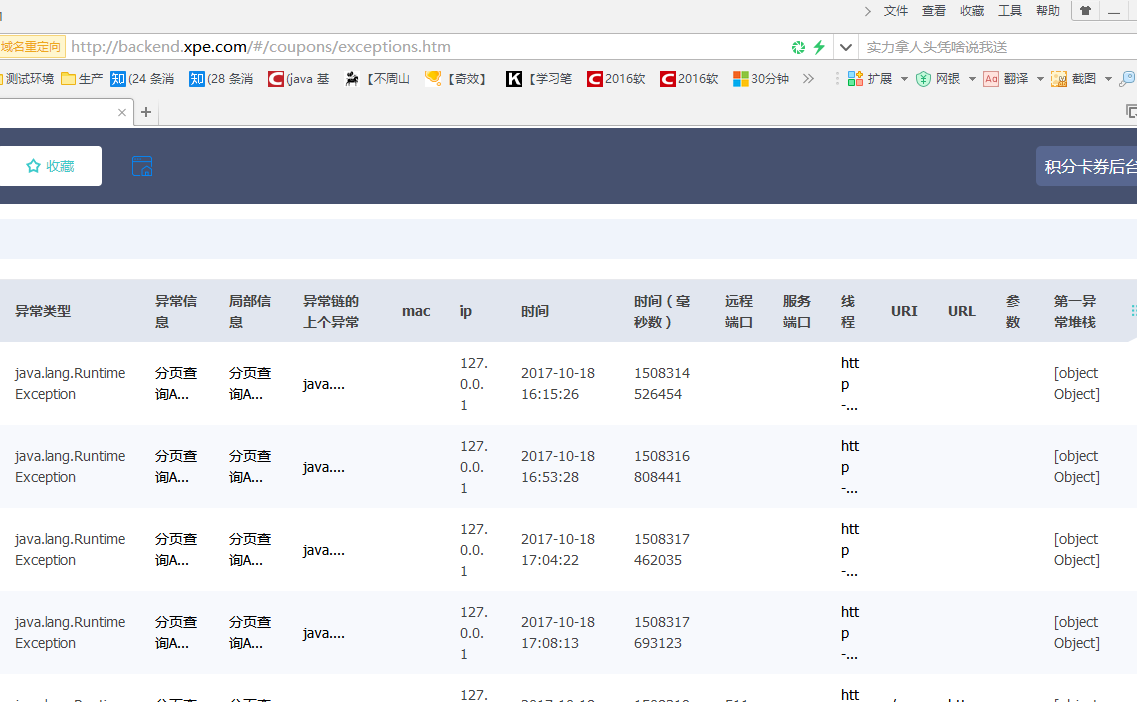
现在是要kinaba进行查询：GET /myexceptions/exception/\_search





说明数据存储成功。

也可以从elasticsearch中将数据查出来：



注：桌面上建立一个文件夹如下：用来存放flume的数据和kafka的日志，配置文件可配：

