有赞使用ES搭建日志平台的演化

raorong@youzan.com

日志平台

实时采集

支持不同语言(Java,PHP,Go,Node)业务系统上报日志的实时收集

实时消费

保证日志产生到消费的实时性，对接流计算平台进行实时消费

日志管理/查询

将不同应用、不同系统、机器等来源系统数据中心化，快速定位系统或应用的问题，提高运维效率

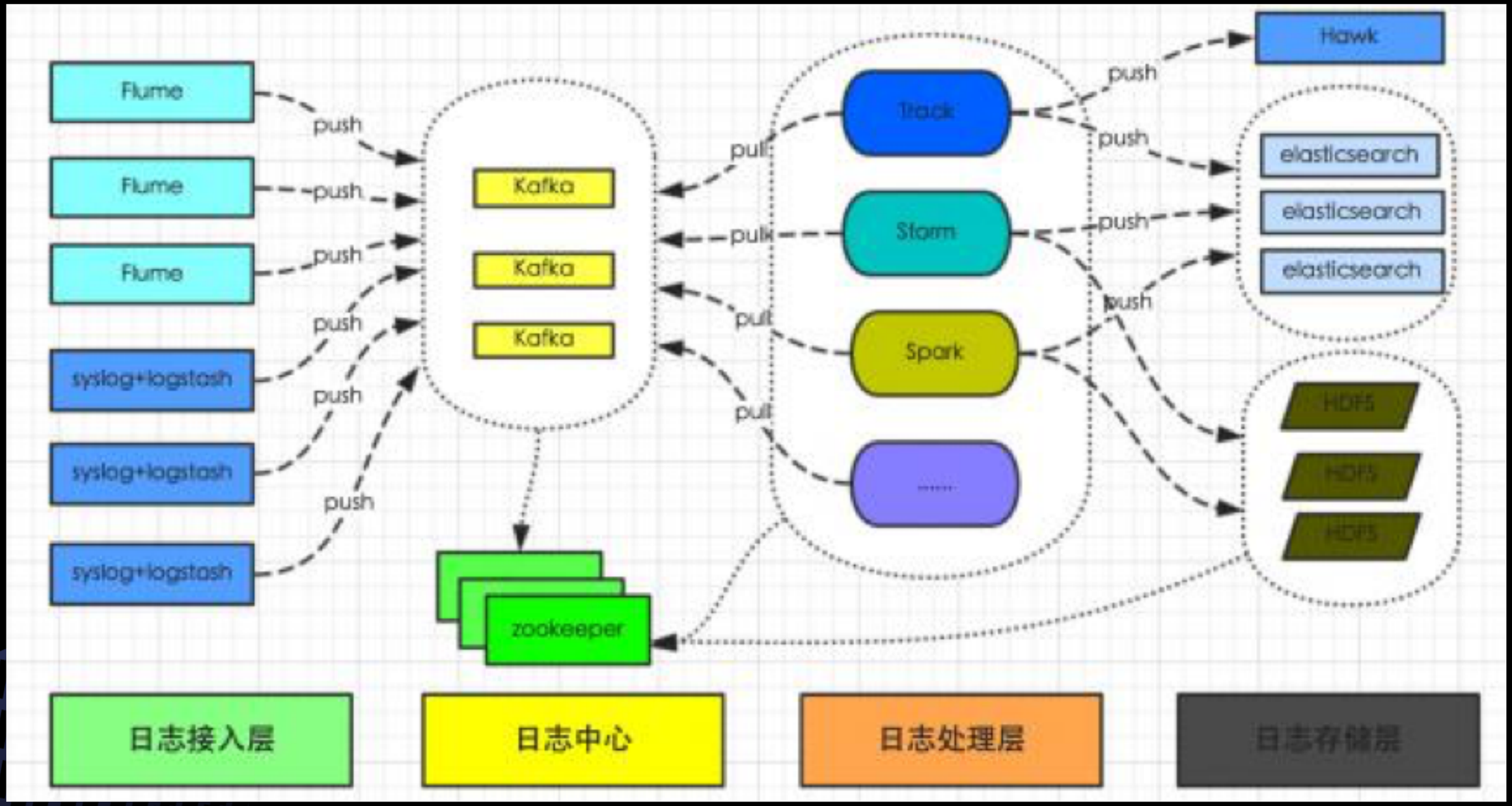
实时报警

异常日志实时通知业务方干系人员，及时解决线上突发状况

CONTENTS

- 日志系统1.0
  - 模块分解
- 日志系统2.0
  - 日志协议
  - 日志采集端优化
- 日志系统3.0
  - Es运维管理
- 总结/展望

日志系统1.0


Flume
Flume
Flume
syslog+logstash
syslog+logstash
syslog+logstash
push
Kafka
Kafka
Kafka
zookeeper
pull
Track
Storm
Spark
.....
Hawk
elasticsearch
elasticsearch
elasticsearch
HDFS
HDFS
HDFS
日志接入层
日志中心
日志处理层
日志存储层

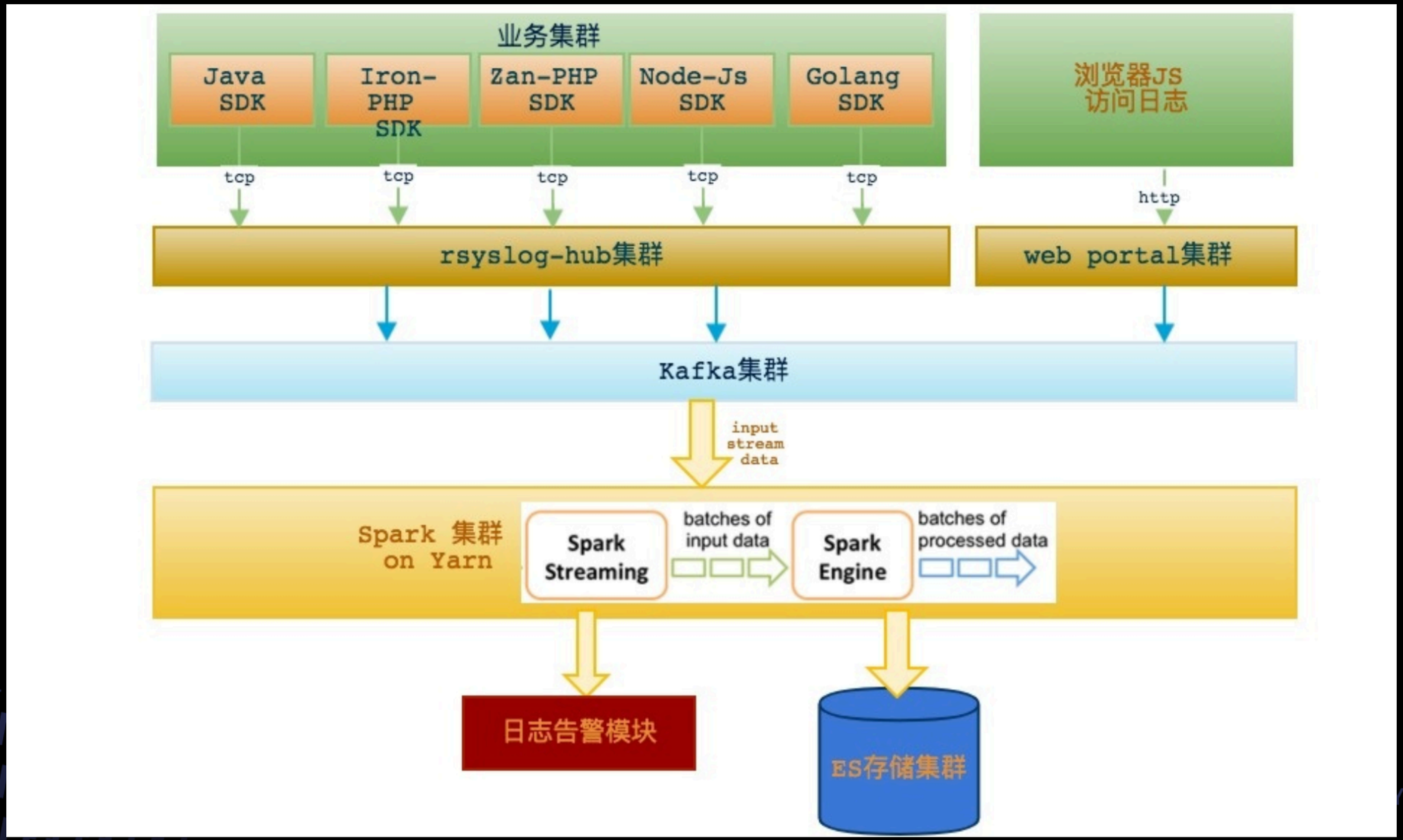
模块分解

- 日志转发
  - 基于rsyslog和logstash
  - 基于flume-ng，只使用了Agent层（包含Source，Channel和Sink，三者组建了一个Agent）
- 日志缓冲层
  - 日志平台使用kafka集群作为自己的缓冲层，为后面的分布式日志消费服务提供异步解耦、削峰填谷的能力，也同时具备了海量数据堆积、高吞吐读写的特性。
- 日志处理层
  - 使用spark streaming流计算框架来消费写入kafka的业务日志
- 日志存储
  - 写入到SSD盘的ES集群

遇到的问题

- 日志样式多
  - 日志格式不统一，新应用接入需要重新开发消费模块，成本过高
  - 多种采集方式，运维困难
- ES 索引内存占用高，索引的管理与维护困难

日志系统2.0



索引设计-日志协议


```
{
  "app":"track",
  "detail":{
    "extra":[
      {
        "aaa":123,
        "bbb":"ccc"
      }
    ],
    "error":{
      "code":500,
      "file":"LogTest.java",
      "line":16,
      "message":"/ by zero",
      "stacktraces":"java.lang.ArithmeticException:/ by zero at com.youzan.paas.track.LogTest.main(LogTest.java:16) "
    },
    "logStatisticsField":"t_track_indexAwarnAtest test rr test.AshowAtrackA111001"
  },
  "level":"error",
  "module":"t_track_index_logger",
  "platform":"JAVA-1.8.0_92",
  "tag":"2017-08-29 18:22:14 - 计算发生异常"
}
```

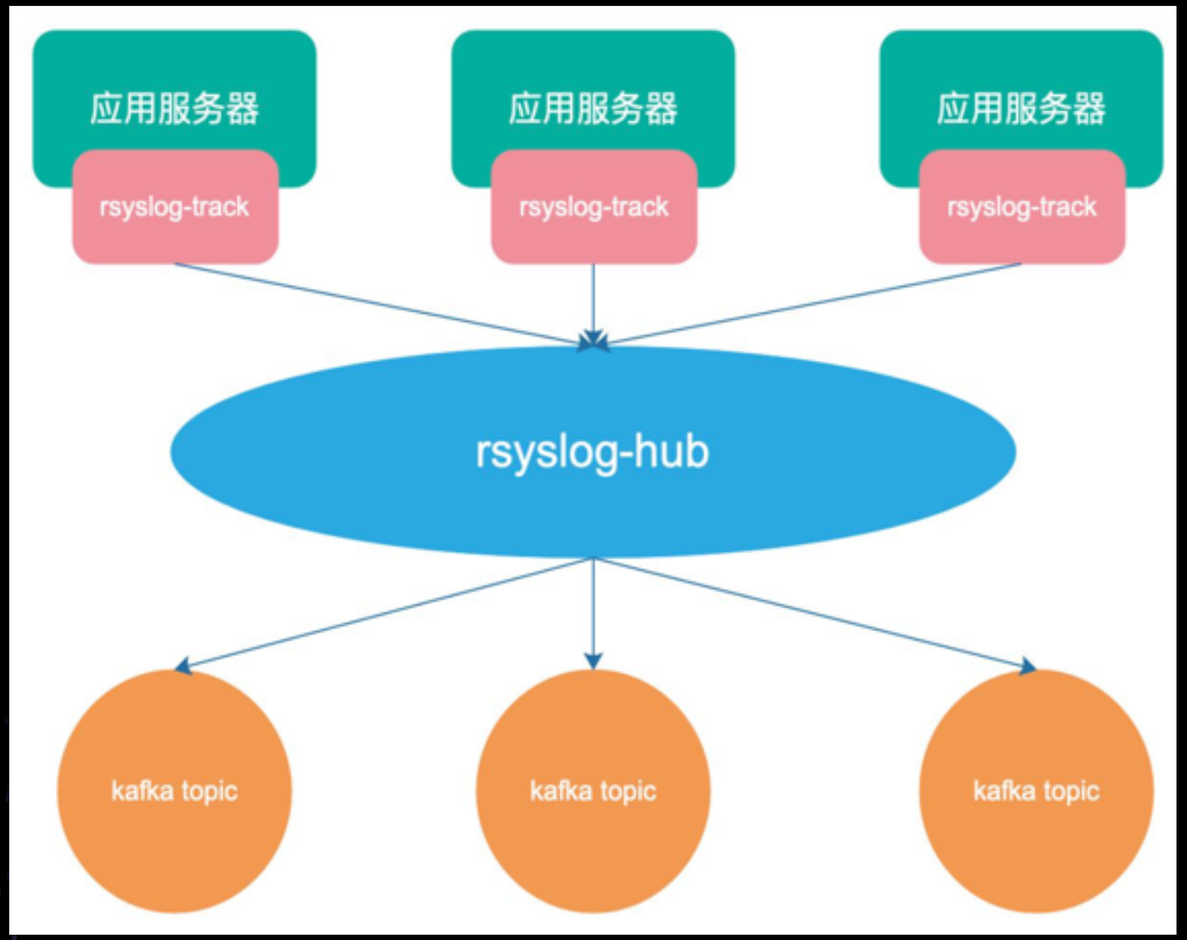

索引字段	字段类型
actionType	string
bizAmount	long
bizNo	string
bizSubNo	string
createTime	date
feeAmount	long
feeAmountSubsidy	long
feeType	integer
feeType	integer
intOut	integer
modeType	string
partnerId	string
prodType	string

字段名称

字段类型

日志采集端

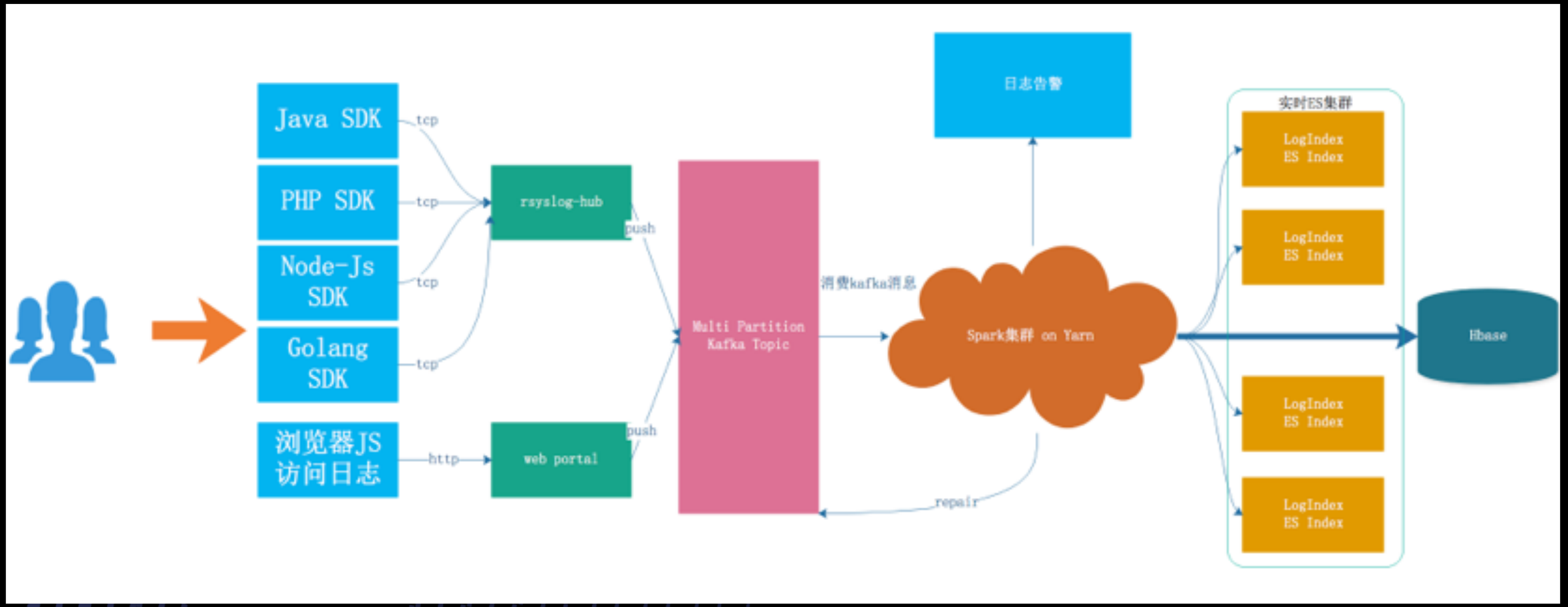
- flume —> rsyslog
- Rsyslog-track转发
- Rsyslog-hub统一发送到kafka



遇到的问题

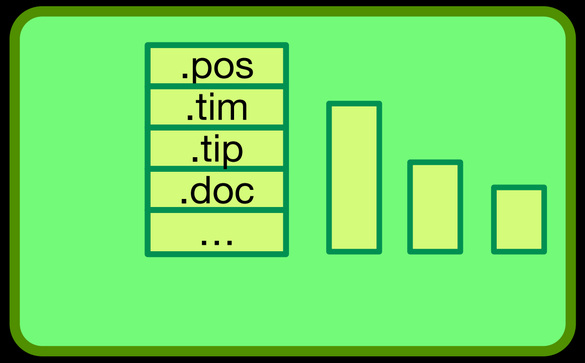
- Es存储成本高
  - 磁盘使用率单机高达70%~80%
- 随着业务的增长，数据倾斜
  - 某一个分片上的执行时间明显高于其他分片，毛刺严重
  - 存储容量不均衡，热点问题
- 数据可能丢失

日志系统3.0



索引设计-基本元素

- 段：segment，索引的基本工作单位
- Lucene索引：段文件集
- 分片：等同于一个Lucene索引
- Elasticsearch索引：若干分片的数据集
- 主从副本：是独立的Lucene索引，ES来控制数据双写



索引设计-索引隔离

- 分片隔离
  - 指定分片隔离的数量，请求发送至全部分片，聚合各分片数据后返回
- 索引隔离
  - 根据指定的策略计算候选节点，指定属性的若干个es node
  - 限制单个节点的分片数量
  - 同一分片主从副本不能在同一个host中出现
  - 同一个index 主分片不能在同一个node节点上
- 集群隔离
  - 按照不同的业务场景，或者核心级别隔离集群

索引设计

- 天网申请的logIndex直接对应ES里的索引结构
- 按照设置的时间周期创建索引（默认是天）
- 自持自定义字段
- 默认只对日志内容字段分词
- 根据写入量动态调整分片数量

```
{
  "log-skynet-log-buyaodongplease-20190422":
{
    "mappings": {
      "log-skynet-log-buyaodongplease": {
        "dynamic": "strict",
        "_all": {
          "enabled": false
        },
        "_source": {
          "enabled": false
        },
        "properties": {
          "env": {
            "type": "keyword"
          },
          "hostname": {
            "type": "keyword"
          },
          "level": {
            "type": "keyword"
          },
          "name": {
            "type": "keyword"
          },
          "tag": {
            "type": "text",
            "analyzer": "ik_max_word"
          },
          "time": {
            "type": "date",
            "format": "yyyy-MM-dd HH:mm:ss"
          }
        }
      }
    }
  }
}
```

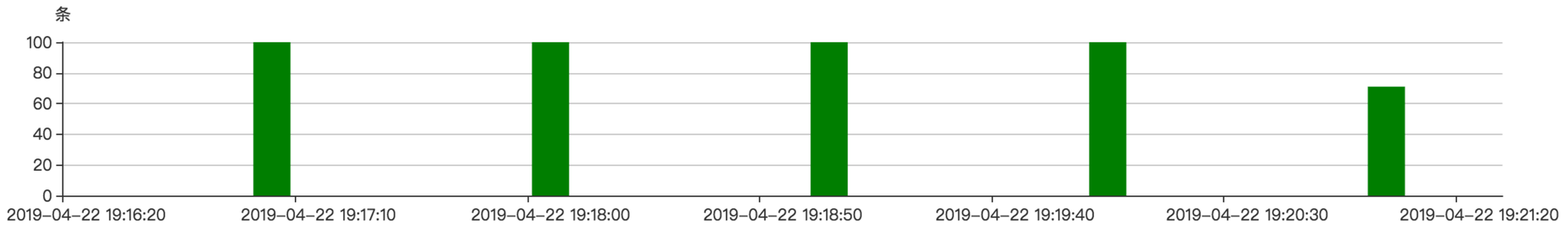
有赞天网监控 Beta　仪表盘　搜索　Zan-Alert　设置　饶榕

默认

应用名称　logIndex　排序选择(降序)　日志内容

skynet-log　buyaodongplease　相关性　时间　请输入日志内容　搜索　保存

全部主机　自定义索引字段精确匹配　5分钟　30分钟　1小时　12小时　1天　2019-04-22 19:17:14 - 2019-04-22 19:21:24



1 2 3 4 5 6 ··· 24　20 条/页　共 471 条　全部　DEBUG　INFO　ERROR　导出　日志量大盘

日志概况　2019-04-22 19:17:05 qabb-qa-skynet2/10.9.120.246 info　展开上下文

日志内容　buyaodongplease, test log!!

日志概况　2019-04-22 19:17:05 qabb-qa-skynet2/10.9.120.246 info　展开上下文

日志内容　buyaodongplease, test log!!

YOUZAN power