

同程旅行千亿级应用日志平台架构演进

周祝群 同程旅行-研发中心







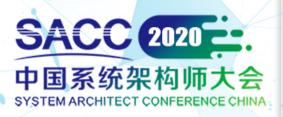
01.目标

02.架构演

03.挑战与收益

04.未来展望

CONTENTS



Part1:目标





日志规模



1600亿

条数/天

280万条/秒

300TB

原始大小/天





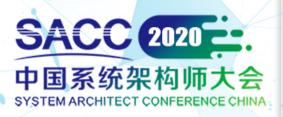
目标

SACC 2020 :- .
中国系统架构师大会
SYSTEM ARCHITECT CONFERENCE CHINA

- 1. 快-延迟低,查询快
- 2. 高可用-可用率高,扩展性好
- 3. 功能强大-不仅仅是个日志查询平台
- 4. 性价比高



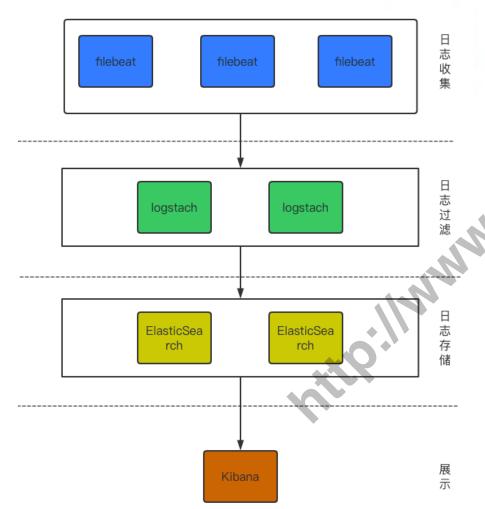




Part2:架构演进









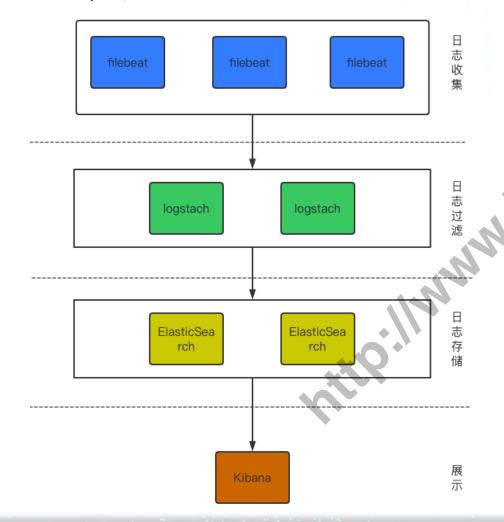
ELK

简单







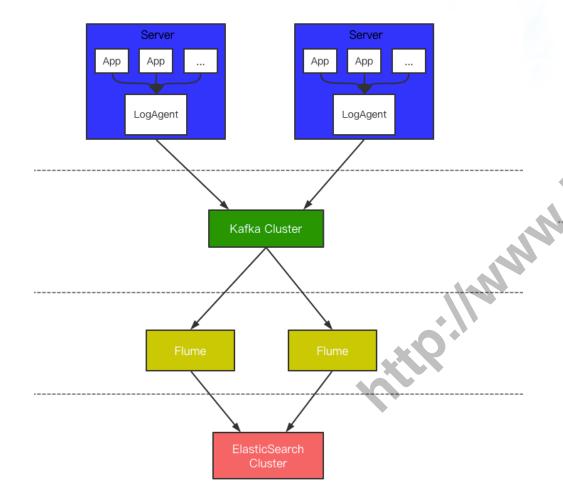




- 1.Logstach性能以及扩展性不 太好
- 2.Es故障时影响较大







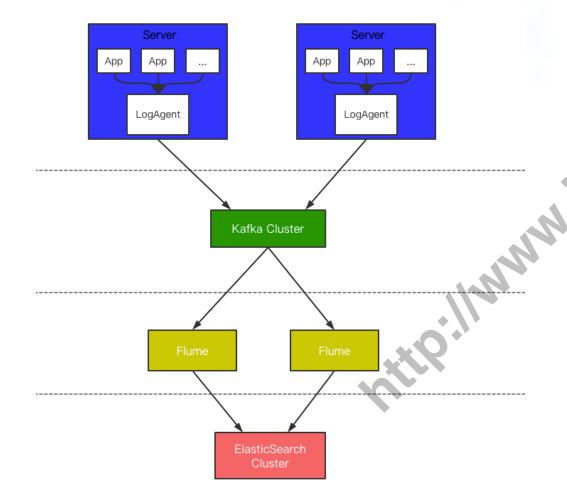


- 1.扩展了filebeat,支持thrift接口写入数据
- 2.引入kafka
- 3. Logstach改为flume

- 1.采集端与存储端解耦,可以互不影响
- 2.支持非文件采集的日志收集
- 3.引入了更轻量且扩展性、性能更好的flume





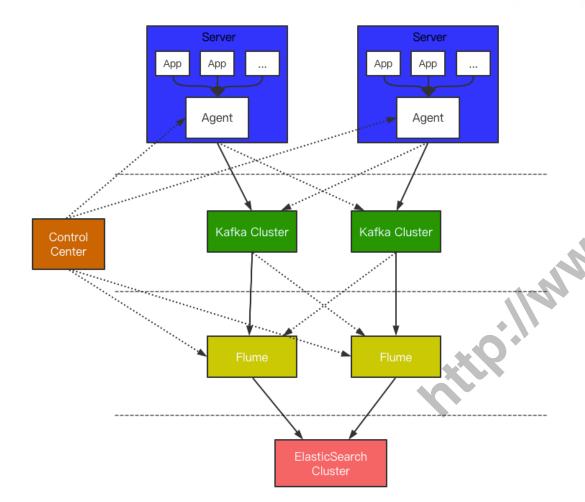




- 1.Flume持续full gc假死
- 2.Kafka单点,Kafka无响应 <u>bug</u>









1.Flume 假死 新增dualSelector Memory Channel与F

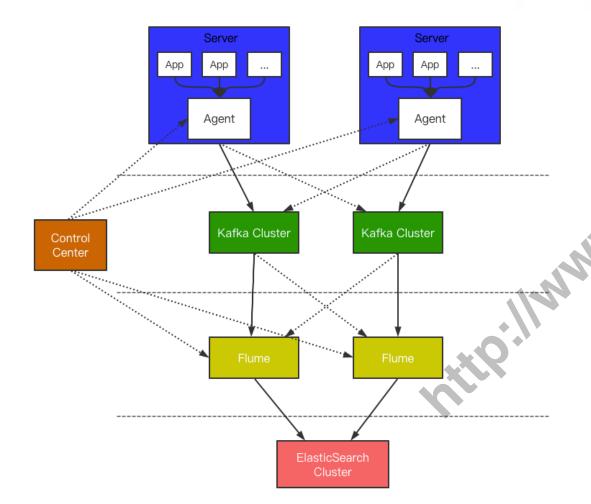
新增dualSelector. Memory Channel与File Channel 自动切换 修改了KafkaSource,支持自动上下线,能者多劳

2.Kafka单点 0.10.1.0 -> 0.10.1.1 <u>bug</u> 部署多套kafka集群

3.新增Control Center。 控制agent 发送的kafka地址 控制flume消费的kafka地址









Elasticsearch 单点

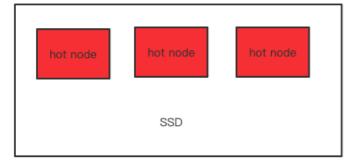


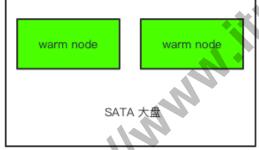


架构演进-v2-单es集群部署



ElasticSearch Cluster





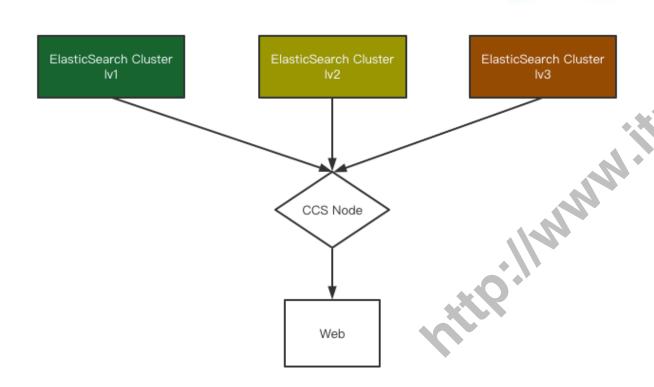
Hot节点,SSD 高配,支撑实时数据

Warm节点,大盘低配机器,存储历史数据









90%的项目占比日志条数 15%

Lv1集群,15%的总量,小规模,高保障,性能强

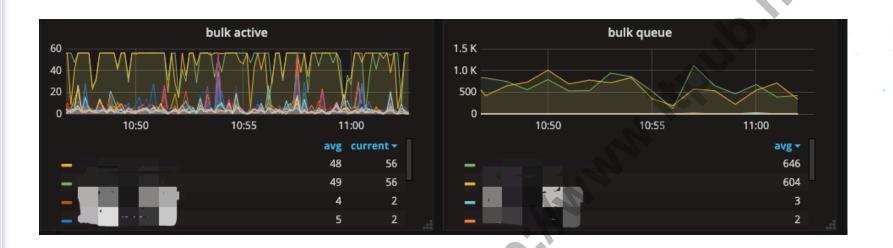
Lv2集群,85%的总量,大规模,高保障,性能优

Lv3集群,X%的总量,warm节点,实时要求低,性能一般









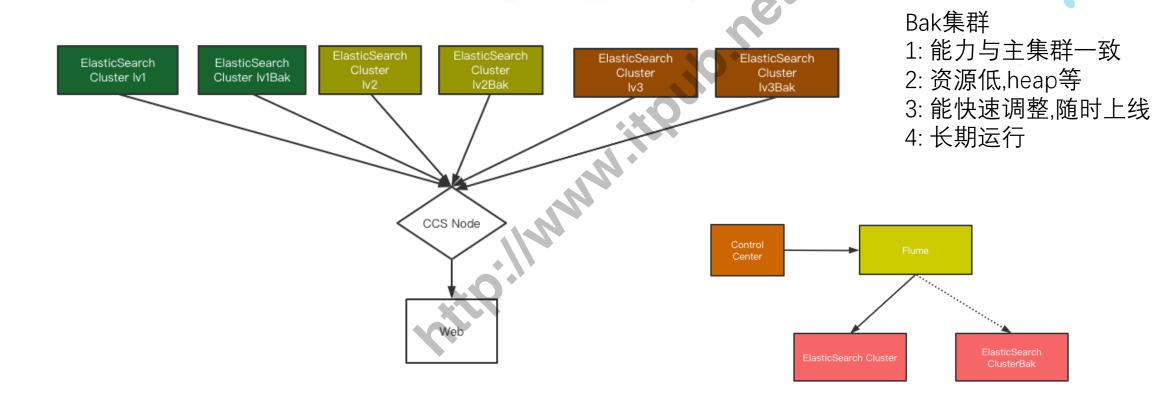
长尾等等原因,集群响应慢。

影响所有写入与查询













架构演进-总结



所有流量都可以动态切换,无任何单点









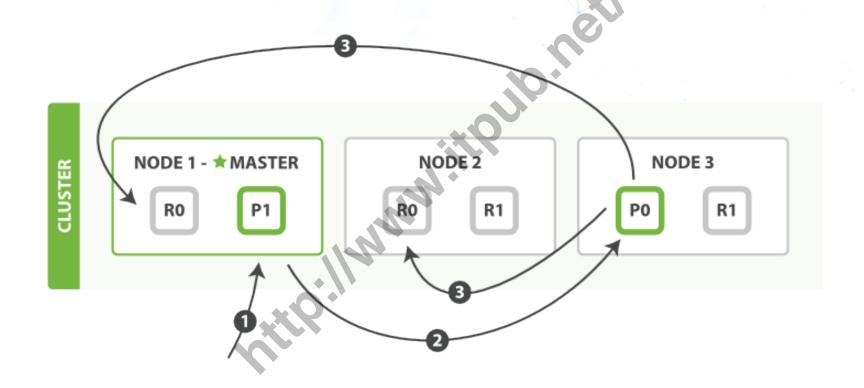
Part3:挑战与收益





挑战-性能与成本



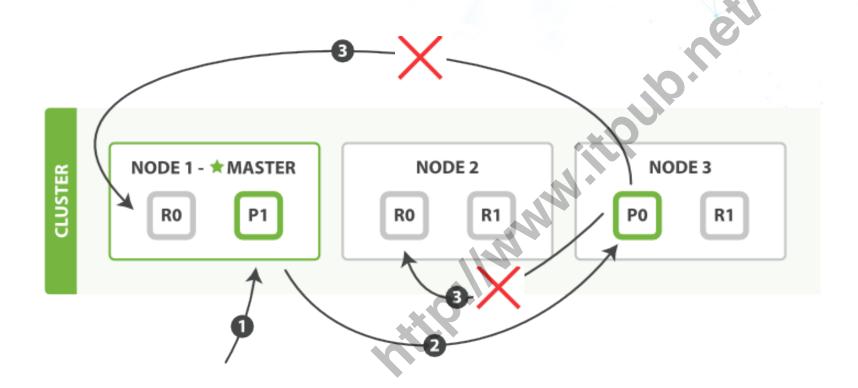






挑战-性能与成本-收益





number_of_replicas:0

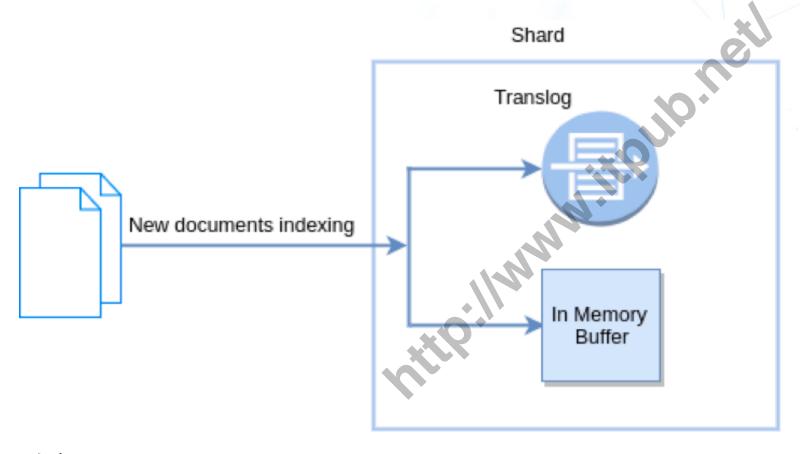
性能提升 100%+





挑战-性能与成本





Tranlog的作用

- 1.恢复时数据回放
- 2.Realtime get



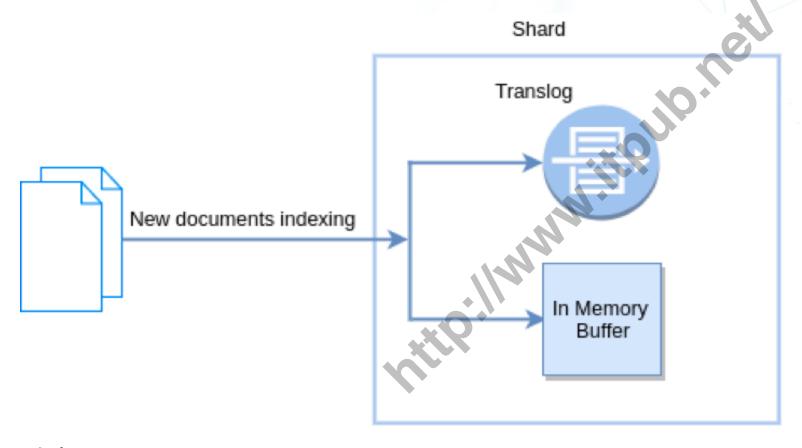
图片来源: https://qbox.io/blog/refresh-flush-operations-elasticsearch-guide





挑战-性能与成本-收益





80%性能提升

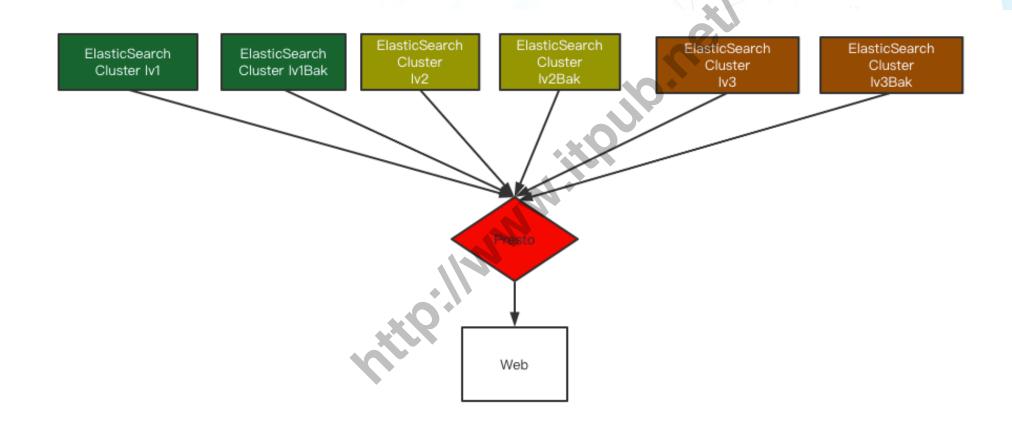
图片来源: https://qbox.io/blog/refresh-flush-operations-elasticsearch-guide





挑战-Es在AP能力弱









挑战-Es在AP能力弱



- 1. 获取原信息慢
- 2. Shard级别的并发,es集群压力较大

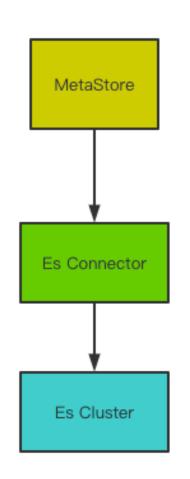






挑战-Es在AP能力弱





MetaStore: 存储es源信息,如shard信息等

Node级别的并发+限速

本地扫描

谓词下推

部分limit下推

...





挑战-Es在AP能力弱-收益



```
select split(split(msg,'serverEndpoint:')[2],', mgr')[1],count(1) from skynet_all_102781 where priority = 2 and dt = '20200916' and logtime > '2020-09-16 11:20:00.000' and logtime < '2020-09-16 12:40:00.000' and msg like '%Redis.RedisTimeoutException%' and ip = '' group by split(split(msg,'serverEndpoint:')[2],', mgr')[1]

SELECT filter1,count(1) from (

SELECT filter1,"json_size"(split(msg,'返回结果:')[2], '$.modules') from skynet_all_103421 where dt = '20181009' and hour = '12' and sub_category='output' and category='heart' and module = 'warden' and msg not like '%wisp.exe%' group by filter1,"json_size"(split(msg,'返回结果:')[2], '$.modules')
) t GROUP by filter1 HAVING count(1) > 1
```

支持实时在线做数据的分析统计



挑战-日志无上下文



什么情况下触发了ERROR日志

展开 2020-10-09 14:36:20.954 ERROR 172.19.103.199(logicidc_hd1) [skynet] [/api/monitor/rule/detail] [search] [] []输入参数{"appId":102781,"beginTime":"2020-10-09 14:12:01.000","endTime":"2020-10-09 14:32:01.000","endTime":"2020-10-09 14:32:01.000","en 4","cPage":1,"pageSize":50,"startRow":1}\norg.springframework.web.client.ResourceAccessException: I/O error on POST request for "http: .com/api/monitor/rule/detail": Read timed o ut; nested exception is java.net.SocketTimeoutException: Read timed out at org.springframework.web.client.RestTemplate.doExecute(RestTemplate.java:696) at org.springframework.web.client.RestTemplate.execute(RestTemplate.java:654) at org.springframework.web.client.RestTemplate.exchange(RestTemplate.java:574) at com.ly.infra.skyeye.metadata.service.api.skynetAPIService.reqInternal(SkynetAPIService.java:207) at com.ly.infra.skyeye.metadata.service.api.skynet.SkynetAPIService.request(SkynetAPIService.java:165) at com.ly.infra.skyeye.metadata.service.api.skynet.SkynetAPIService.proxySearch(SkynetAPIService.java:136) at com.ly.infra.skyeye.metadata.web.skynet.SkynetController.skynetProxySearch(SkynetController.java:76) at com.ly.infra.skyeye.metadata.web.skynet.SkynetController\$\$Fast@lassBySpringCGLIB\$\$ef13547d.invoke(<generated>) at org.springframework.cglib.proxy.MethodProxy.invoke(MethodProxy.java:204) at org.springframework.aop.framework.CqlibAopProxy\$CqlibMethodInvocation.invokeJoinpoint(CqlibAopProxy.java:746) $at \verb| org.springframework.aop.framework.ReflectiveMethodInvocation.proceed (ReflectiveMethodInvocation.java: 163) \\$ at com.elong.hotel.core.ChecklistLogAspect.invoke(ChecklistLogAspect.java:42) at org.springframework.aop.framework.ReflectiveMethodInvocation.proceed(ReflectiveMethodInvocation.java:185) at org.springframework.aop.interceptor.ExposeInvocationInterceptor.invoke(ExposeInvocationInterceptor.java:92) at org.springframework.aop.framework.ReflectiveMethodInvocation.proceed(ReflectiveMethodInvocation.java:185) at org.springframework.aop.framework.cqlibAopProxy\$DynamicAdvisedInterceptor.intercept(CglibAopProxy.java:688) at com.ly.infra.skyeye.metadata.web.skynet.SkynetController\$\$EnhancerBySpringCGLIB\$\$6c373214.SkynetProxySearch(<generated>) at sun.reflect.GeneratedMethodAccessor140.invoke(Unknown Source) at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43) at java.lang.reflect.Method.invoke(Method.java:498) at org.springframework.web.method.support.InvocableHandlerMethod.doInvoke(InvocableHandlerMethod.java:209) at org.springframework.web.method.support.InvocableHandlerMethod.invokeForRequest(InvocableHandlerMethod.iava:136) g.springframework.web.servlet.myc.method.annotation.ServletTnyocableHandlerMethod.inyokeAndHandle(ServletTnyocableHandlerMethod.iaya:102)







挑战-日志无上下文



借助Java Agent,实现日志方法拦截 注入MDC 解析外部系统的mapping,如域名日志与traceld关系







挑战-日志无上下文



日志链路 调用链路 2020-10-09 14:34:45.083	INF [Time] [fastQuery] [枸建结果] [] []耗时62ms
	INF
	INFO 基于分钟查询] [es获取] [] [goods-service-nb::com.elong.hotel.goods.goodservice.utils.StaticDetailOperUtil::batchGetAllStaticList] [2020-10-09 13:34:00]{"dataKey":null,"logTime":"2020-1 -09 13:34:00", minute":"","day":"","month":"","iserverMame":"","serverGroup":"","secuteCount":3428, "executeCountColor":"","exceptionRate":0.0, "exceptionRateColor":"","bizExceptionRateColor":"","bizExceptionRateColor":"","bizExceptionRateColor":"","bizExceptionRateColor":","","bizExceptionRateColor":","","bizExceptionCount":0,"computeNode":null,"totalTimeSpan": 0.0,"totalTime":47372.0,"key":"2020-10-09 13:34:00","sumbizExceptionCount":{"value":0},"sumExceptionCount":{"value":3428},"sumTotalTimeSpan":{"value":47372},"maxExecuteTimeSpanA 1g":{"value":138},"avgTop99ExecuteTime":{},"time":"2020-10-09 13:34:00","top99ExecuteTime":0.0}
日志链路 调用链路 2020-10-09 14:34:45.074	INFO 基于分钟查询] [skyeye默认新建] [] [goods-service-nb:;com.elong.hotel.goods.goodservice.impl.ThriftProxyImpl::searchHotel] [2020-10-09 14:34:00]skyeye默认新建
日志链路 调用链路 2020-10-09 14:34:45.074	INFO [基于分钟查询] [skyeye默认新建] [] [goods-service-nb::com.elong.hotel.goods.goodservice.impl.ThriftProxyImpl::searchHotel] [2020-10-02 14:34:00]skyeye默认新建
	INFO [基于分钟查询] [es获取] [] [goods-service-nb::com.elong.hotel.goods.goodservice.impl.ThriftProxyImpl::setMajiaRequest] [2020-10-02 14:34:00]{"dataKey":null,"logTime":"2020-10-02 14:34:00","minute":"","day":"","month":"","idc":"","serverName":"","serverGroup";"","exceptionRateColor":"","exceptionRate":0.0,"exceptionRateColor":"","bizExceptionRate":0.0,"bizExceptionRateColor":"","widexecuteTimespan":20.444,"avgExecuteTimespanColor";"","maxExecuteTimespan":77,"maxExecuteTimespanColor":"","exceptionCount":0,"bizExceptionCount":0,"computeNode":null,"totalTimespan":0.0,"totalTimespan":178480.0,"key":"2020-10-02 14:34:00","sumbizExceptionCount":{"value":0},"sumExecuteCount":{"value":8730},"sumTotalTimeSpan":{"value":178480},"maxExecuteTimeSpanAgg":{"value":77},"avgTop99ExecuteTime":{},"time":"2020-10-02 14:34:00","top99ExecuteTime":0.0}
	INFO 基于分钟查询] [gods-service-nb::com.elong.hotel.goods.goodservice.impl.ThriftProxyImpl::setMajiaRequest] [2020-10-09 14:33:00]{"dataKey":null,"logTime":"2020-10-09 14:33:00","minute":"","aay":"","month":"","idc":"","serverName":"","serverGroup":"","exceptionRateColor":"","exceptionRate":0.0,"exceptionRate":0.0,"bizExceptionRateColor":"","bizExceptionRate":0.0,"bizExceptionRateColor":"","bizExceptionRate":0.0,"bizExceptionRateColor":"","bizExceptionRate":0.0,"bizExceptionRate":0.0,"bizExceptionCount":0,"computeNode":null,"totalTimeSpan":0.0,"totalTime":"119380.0,"key":"2020-10-09 14:33:00","sumbizExceptionCount":{"value":0},"sumExecuteCount":{"value":5280},"sumTotalTimeSpan":{"value":119380},"maxExecuteTimeSpanAgg":{"value":49},"avgTop99ExecuteTime":{},"time":"2020-10-09 14:33:00","top9ExecuteTime":0.0}





挑战-日志无上下文-收益



tcbase.java.skynetlogv2	2020-10-09 14:25:09.585	INFO	
tcbase.java.skynetlogv2	2020-10-09 14:25:09.559	INFO	
tcbase.java.skynetlogv2	2020-10-09 14:25:09.559	INFO	用户验证,HJ- to - RCSU WUINID:
tcbase.java.skynetlogv2	2020-10-09 14:25:09.558	INFO	用户验证,用户看· Mail WUINID.
tcbase.java.skynetlogv2	2020-10-09 14:25:09.557	INFO	1204213原工号获取用户权限
tcbase.java.skynetlogv2	2020-10-09 14:25:09.557	INFO	1204213新工号获取用户权限
tcbase.java.skynetlogv2	2020-10-09 14:25:09.556	INFO	skyeye 用户设置 1204213
TLB	2020-10-09 14:25:09.555	INFO	详情 请求ur. ipi/skyeye/logs;耗时=39ms
infrascs.java.skyeye.api	2020-10-09 14:25:09.553	INFO	请求ip10.181.18.22, 172.30.109.11, 172.19.28.155
TLB	2020-10-09 14:25:09.528	INFO	详情 请求ur om/skynet/logs/search/api/skyeye/logs;耗时=80ms
TLB	2020-10-09 14:25:09.370	INFO	详情 请求uɪ ni/proxy/skynet/logs/search/api/skyeye/logs;耗时=247ms







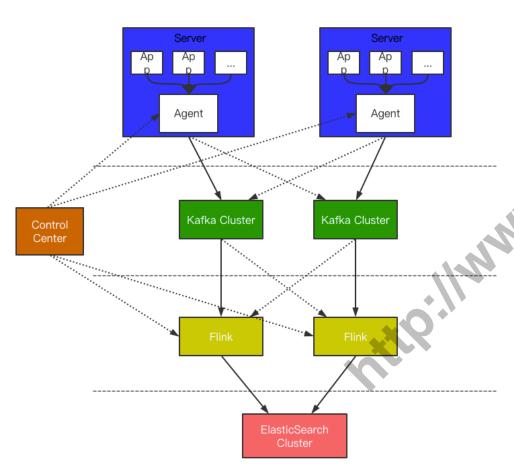
原始数据特别大,但是我仅仅只想做个监控报警













仅需要记录一条日志

Agent识别监控原始数据至独立topic

Flink计算dashboard数据以及报警

Dashboard数据存入es用来展示







D	名称 s1parsenull	字段详情 gmt_create:string,depultyLode:string,arrUityLode:string,depulate:string,retUate:string	操作編辑
7	inbmonitor	k1:string,k2:string,k3:string,k4:string,k5:string,v1:bigint,v2:bigint,tracerid:string,appUk:string	编辑
В	linktracker	traceld:string,contextld:string,memberld:string,serialNo:string,applicationId:string,sceneType:bigint,requestUrt:string,busiModel:bigint,usedTime:bigint,createTime:bigint,responseTime:bigint,isAsyn:bigint,logType:bigint,ex1:string,ex2:string	编辑
9	hotelordercreatesystemmonitorin	Customer Serialld: bigint, Order Serialld: bigint, Order Serialld: bigint, Order Serialld: bigint, ExtrapElong: string, B2CUserld: bigint, StartDate: string, EndDate: string, Hotelld: bigint, Roomld: bigint, Policyld: bigint, IsCtripElong: string, IsOrder Create: string, Fail Codeld: bigint, Fail Message: string, All Amount: bigint, Pay	编辑
10	hotelpaynotifyindex	CustomerSerialId:bigint,OrderSerialId:string,CreateDate:bigint,FallMessage:string,FallCode:bigint,IsSuccess:string,SpecificErrorMessage:string,ext1:string,ext1:string,ext2:string	编辑
11	hotelorderrefundindex	Customer Serial Id: bigint, Order Serial Id: bigint, Order Serial Id: bigint, Order Serial Id: bigint, Refund Type Id: bigin	编辑
12	hotelincrementprocessindex	CustomerSerialId:bigint,OrderSerialId:string,CreateDate:bigint,SourceFrom:bigint,FaillCodeld:bigint,FaillMessage:string,IsSuccess:string,ExternalSerialId:string,SourceFrom:Name:string,SpecificErrorMessage:string,ext1:string,ext2:string	编辑
13	distributionbookablecheck	DistributorId: bigint, Hotelld: bigint, Hotelld: bigint, Foomld: bigint, Foomld: bigint, Cane Date: bigint, Come Date: bigint, Booking Count: bigint, Guest Count: bigint, Cane Booking: bigint, Error Message: string, Source From: bigint, Calendar Time: bigint, Booking Count: bigint, Cane Date: bigint, Cane Date: bigint, Booking Count: bigint, Cane Date: bigint, Cane Date	编辑
14	hotelordersearchtimesindex	CreateDate:blgint,Channel:string,SearchTimes:blgint,Ex:object	编辑
15	hotelorderinsuranceindex	Customer Serial Id: bigint, Order Serial Id: bigint, Order Serial Id: bigint, Buylnsurance From: big	编辑
16	hotelorderswitchcreateindex	Customer Serial Id:string, Order Serial Id:string, Sub Order Serial Id:string, Create Date: bigint, B2CUserId: bigint, Hotelld: bigint, Foolicy Id: bigint, Foolicy Id: bigint, Fail Codeld: bigint, Fail Codeld: bigint, Fail Message: string, Client Type: bigint, Flatform Id: bigint, Specifi	编辑
17	hotelbillapplyindex	CustomerSerialId:string,OrderSerialId:string,CreateDate:bigint,BillType:bigint,InvoiceId:bigint,SpecificErrorMessage:string,IsSuccess;bigint,ex:object	编辑
18	hotelctriporderpushindex	CustomerSerialId:string,OrderSerialId:string,SubOrderSerialId:string,CreateDate:bigint,SourceFrom:bigint,SourceFrom:bigint,SourceFrom:Digint,SourceSerialId:string,IsSuccess:bigint,SpecificErrorMessage:string,ex:object	编辑
20	sjmdffq	dbName:string,tableName:string,timelong:bigint,size:bigint,time:bigint,jobld:string,ext:object	编辑
21	skynetlogmonitor	data:string	编辑
23	hoteladditional	product_kind_code:string,product_code:string,is_flash:bigint,order_amount:string,settlement_type:bigint,time_consuming:bigint,dynamic_proportion:string	編辑
47	esapiindexinfo	clusterId:string,clusterName:string,health:string,index:string,pri:bigint,priRate:double,allRate:double,storeSize:double,time:date,type:string	编辑
19	esapinodeinfo	key:string,index:double,search:double,segmentsMemory:double,heapPercent:double,clusterId:string,clusterName:string,time:date,type:string,segmentsCount:double,host:string	编辑
in	peanithreadnool	kovertrina antiverdavible avecuardavible timerdate tunerstrina alveterida eliveterida alveterida nelveterida alveterida eliveterida alveterida eliveterida elivete	编辑

提前定义好字段











配置好监控报警策略





挑战-监控报警原始数据浪费-收益





100倍+ 存储量减少







Part4:未来展望







未来展望

SACC 2020 主. 中国系统架构师大会 SYSTEM ARCHITECT CONFERENCE CHINA

- 1. 寻找性价比更高的存储解决方案
- 2. 完成更多的基于日志的数据挖掘







