

ES 在百度Aladdin Trace系统的应用

王鹏

内容简介

我们基于elastic（Elasticsearch+Kibana）做了一套针对Aladdin离线建库业务的trace系统，用于解决问题追查、数据统计、指标可视化等业务痛点。

主要内容

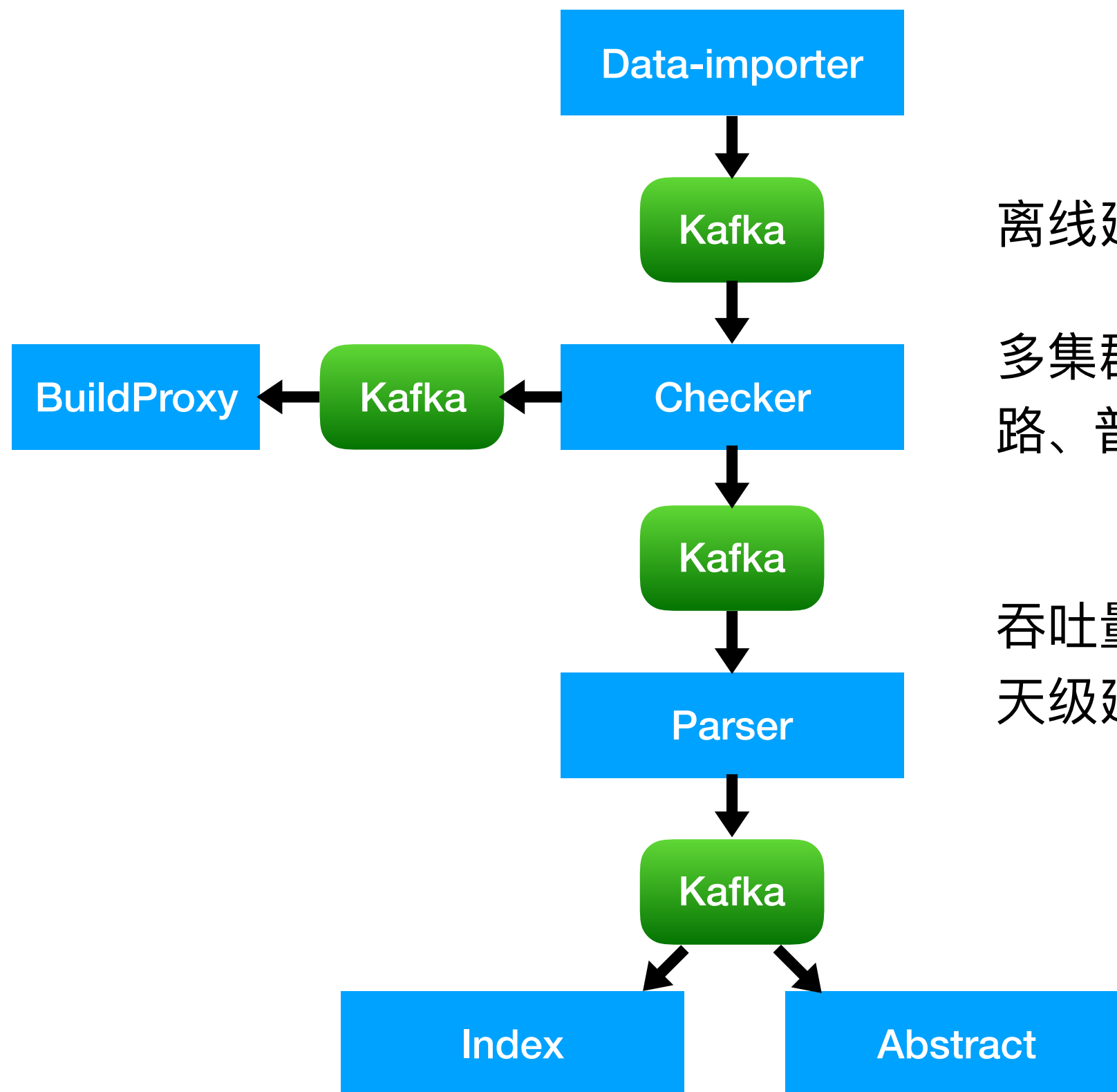
- 背景
- 思考
- 实现
- 效果

背景

aladdin简介：特型卡片，一套检索系统。aladdin离线建库系统简介

存在的问题：aladdin离线模块目前在追查问题、建库数据统计上，存在一些问题，影响case追查、问题分析、指标衡量等。急需一个Trace系统，提升case追查效率，及时反馈问题等。

背景



离线建库系统介绍

多集群异构部署（比如高时效通路、普通通路、批量任务通路等）

吞吐量：入口模块平均5Wqps，
天级建库成功数据量平均8亿条。

思考

理清当前的现状： 离线建库系统架构、数据处理流程、日志保存形式、日志格式、日志量等。

方案选型： trace方案： Google—Dapper, Twitter—zipkin, 国内开源、公司内部。

日志处理、存储、检索等考虑： 首先检索速度必须快、服务易维护、资源占用尽量少、上手简单。

实现

日志传输：公司内的服务，类似kafka，可以实时订阅日志（全机部署，稳定、低延迟）

日志处理：自己写的类似logstash服务，解析日志字段，前提是确定好Elasticsearch的mapping。

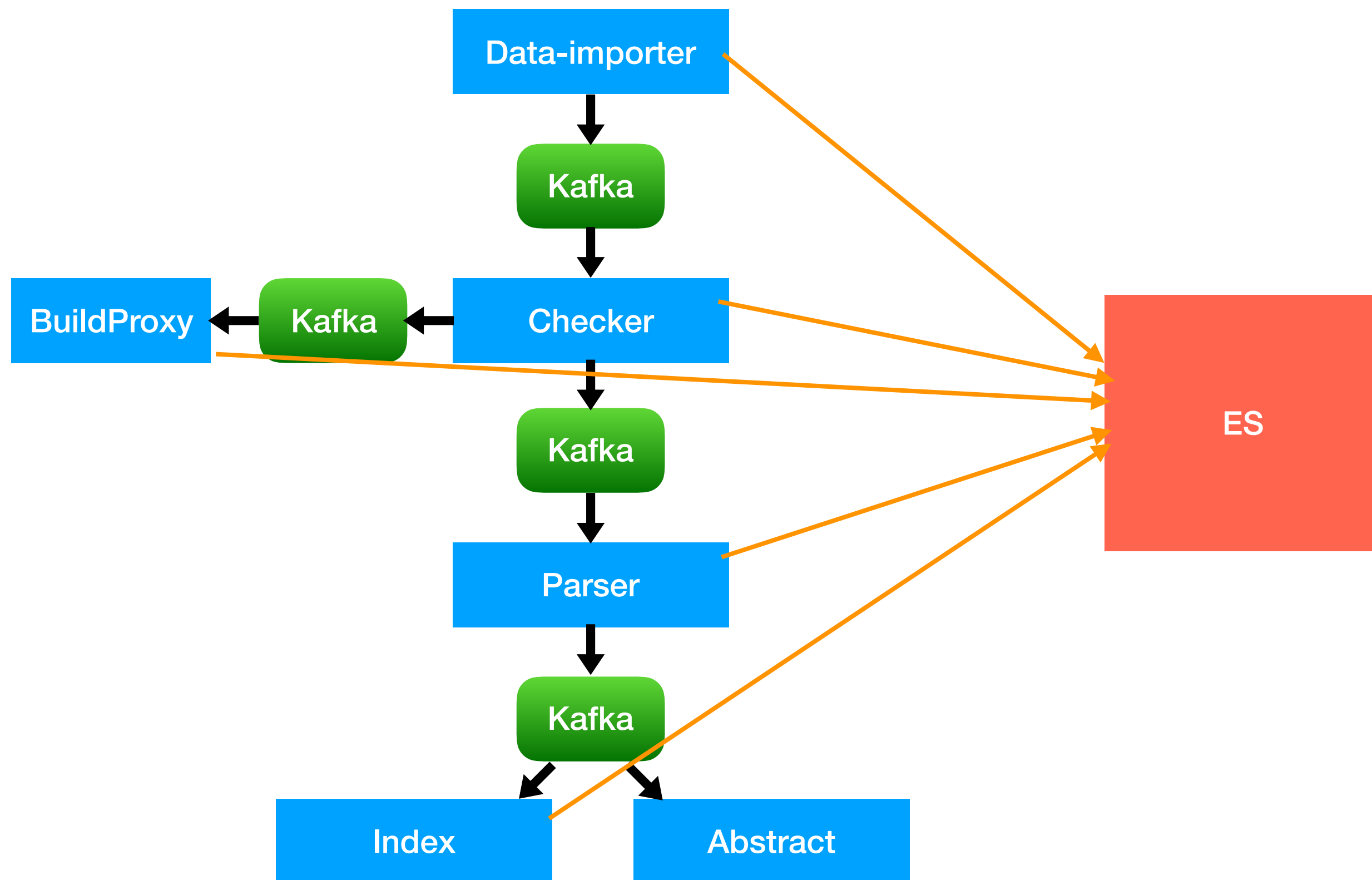
日志存储、检索：Elasticsearch

可视化：Kibana

相关：其中Elasticsearch+Kibana是6.0.0版本，2017.10开始接触，一开始使用的5.4，一个月之后就有6，于是就使用当初最新版。

写ES用的[开源C++实现](#)

实现



实现

分模块建index，使用mapping中的字段区分模块的异构集群，比如checker模块有三种通路：高时效、普通、批量灌库，那么分别用es_type_name: effect_checker, checker, conveyor_checker

Index 使用 servername_data，例如test_build_parser_20180909

日志保存3天，CT任务每天凌晨执行：删除3天之前的index，新建后两天的index

实现

```
#!/bin/bash
```

```
new_data=`date -d "-1 days ago" +%Y%m%d`
```

```
curl -XPUT "http://atrace-all.nlp.yq01.serv:8810/test_build_check_$new_data?pretty" -H 'Content-Type: application/json' -d'
```

```
{  
  "settings": {  
    "number_of_shards" : 128,  
    "index.translog.flush_threshold_size" : "1024mb",  
    "index.translog.durability" : "async",  
    "refresh_interval" : "30s",  
    "number_of_replicas": 0  
  },  
  ...
```

实现

```
"mappings": {  
  "check" : {  
    "properties" : {  
      "log_server_name" : {  
        "type" : "text"  
      },  
      "log_module_name" : {  
        "type" : "text"  
      },  
      "log_full_msg" : {  
        "type" : "text",  
        "index": "false"  
      },  
      "log_feature_name" : {  
        "type" : "keyword"  
      },  
      "log_inside_date" : {  
        "type" : "date",  
        "format": "MM-dd"  
      },  
      "log_inside_time" : {
```

```
"versionid" : {  
  "type" : "long"  
},  
"logid" : {  
  "type" : "keyword"  
},  
"loc" : {  
  "type" : "keyword"  
},  
"priority" : {  
  "type" : "keyword"  
},  
"errid" : {  
  "type" : "keyword"  
},  
"send2psr" : {  
  "type" : "keyword"  
},  
"send2proxy" : {  
  "type" : "keyword"  
},
```

效果

ip	heap.percent	ram.percent	cpu	load_1m	load_5m	load_15m	node.role	master	name
10.94.223.212	41	99	25	1.65	1.44	1.73	mdi	-	node-19
10.94.223.210	71	99	23	1.98	1.56	1.33	mdi	*	node-15
10.99.243.211	41	99	25	0.86	1.06	1.17	mdi	-	node-44
10.99.206.24	43	100	22	3.60	3.43	2.94	mdi	-	node-11
10.99.207.12	53	99	16	1.44	1.08	1.00	mdi	-	node-13
10.86.55.44	68	98	6	0.26	0.27	0.25	mdi	-	node-32
10.94.208.159	24	99	51	6.17	5.85	5.65	mdi	-	node-18
10.99.200.155	61	99	24	2.38	1.65	1.52	mdi	-	node-4
10.99.193.230	59	99	20	1.51	1.24	1.27	mdi	-	node-41
10.99.192.76	54	99	21	0.68	0.93	1.15	mdi	-	node-3
10.94.215.232	64	99	27	0.86	1.19	1.29	mdi	-	node-45
10.94.209.212	13	99	12	0.12	0.36	0.42	mdi	-	node-24
10.86.55.28	60	97	6	0.35	0.28	0.21	mdi	-	node-31
10.94.208.16	11	99	55	4.63	4.10	3.80	mdi	-	node-21
10.99.202.232	57	99	18	0.90	1.02	1.18	mdi	-	node-9
10.94.213.243	48	99	18	0.78	0.96	1.17	mdi	-	node-43
10.94.243.113	69	99	26	1.23	1.16	1.27	mdi	-	node-17
10.86.64.33	9	98	5	0.08	0.09	0.18	mdi	-	node-30
10.94.219.199	54	99	24	1.33	1.69	2.16	mdi	-	node-40
10.99.194.78	22	99	20	2.18	2.34	2.35	mdi	-	node-5
10.94.239.242	59	99	35	1.12	1.74	2.04	mdi	-	node-14
10.101.46.49	65	99	34	3.26	2.97	2.94	mdi	-	node-26
10.95.178.147	20	99	14	0.75	0.87	0.99	mdi	-	node-16
10.100.122.140	40	99	16	0.71	0.94	0.98	mdi	-	node-8
10.94.223.213	13	99	29	1.60	1.72	1.65	mdi	-	node-20
10.101.44.66	51	99	20	1.02	1.08	1.10	mdi	-	node-42
10.99.202.218	26	99	14	1.20	0.89	0.90	mdi	-	node-6
10.99.207.114	46	99	15	0.49	0.76	0.97	mdi	-	node-12
10.86.61.11	11	98	7	2.70	2.65	2.47	mdi	-	node-34

效果

health	status	index	uuid	pri	rep	docs.count	docs.deleted	store.size	pri.store.size
green	open	test_build_dataproxy_jx_20180906	vd6npmJbS62oVokHmCqyGA	128	0	279956308	0	147.4gb	147.4gb
green	open	test_build_as_20180909	ov1lnO79QLu0f8bFELNkCw	128	0	0	0	29.1kb	29.1kb
green	open	test_build_dgw_20180908	6wzjjTODRUuDkrjut4jTIA	64	0	268368177	0	79.9gb	79.9gb
green	open	test_build_as_20180906	9vuS22iLTaul2_s0HA4t9Q	128	0	325801362	0	196.9gb	196.9gb
green	open	test_build_check_20180909	2-AIScC4Tl6n7Qs8g2dZPA	128	0	0	0	29.1kb	29.1kb
green	open	test_build_mip_20180909	Do_cNtyPRIei94VMUjwCDQ	64	0	0	0	14.5kb	14.5kb
green	open	test_build_bp_20180909	0gPxxLrxQ5Gao7ETNHY1BA	96	0	0	0	21.8kb	21.8kb
green	open	test_build_vc_20180909	igtjl0a8TmyH4Uql09968A	36	0	0	0	8.1kb	8.1kb
green	open	test_build_vc_20180906	Z4_Q-5FASS63vnoDV_oKJA	36	0	3503935	0	1.5gb	1.5gb
green	open	test_build_vc_20180910	lhp6IWuyRXKVBFj-esU2wg	36	0	0	0	8.1kb	8.1kb
green	open	test_build_mip_20180906	AOhwwbJeSB20Vg5wTxI9yg	64	0	1135426	0	302.3mb	302.3mb
green	open	test_build_bp_20180906	mz2m8b-FSzwR00Hb4whJA	96	0	804317210	0	264.1gb	264.1gb
green	open	test_build_dgw_20180906	qwfOrpx6Tx2-Rg9Elu-g3w	64	0	337964393	0	97.2gb	97.2gb
green	open	test_build_as_20180908	84wU97shQxW5WOH019U-QQ	128	0	180960324	0	138.8gb	138.8gb
green	open	test_build_bp_20180907	cxaGqb-URP-0NaXdpPXNWw	96	0	727743043	0	237.5gb	237.5gb
green	open	test_build_parser_20180909	Vlcw212BRJ-v5_vsUJ22eA	128	0	0	0	29.1kb	29.1kb
green	open	test_build_mip_20180907	JlQ8mLwpQ6W1QU5ONxMnzA	64	0	1048314	0	279.5mb	279.5mb
green	open	test_build_dataproxy_jx_20180909	KWjrpJ_ESnyFvx8GLtBmSA	128	0	0	0	29.1kb	29.1kb
green	open	test_build_parser_20180906	acHhsTWLQY21rYHq53VEDw	128	0	960151016	0	394.7gb	394.7gb
green	open	test_build_check_20180910	WG7IVJLqTRW_M5annhEgpw	128	0	0	0	29.1kb	29.1kb
green	open	test_build_as_20180910	gz12tREWRyezeWSpk9XNyQ	128	0	0	0	29.1kb	29.1kb
green	open	test_build_dataproxy_jx_20180908	2iwHmE-tSduLSDZpHmBG_Q	128	0	132276838	0	70.7gb	70.7gb
green	open	test_build_as_20180907	sbcgBIk0SPWYh8YxN6q5Lg	128	0	805784335	0	535.2gb	535.2gb
green	open	test_build_dgw_20180907	gjEkvcbpTyadV5_UHldZAw	64	0	548134080	0	155.6gb	155.6gb
green	open	test_build_mip_20180910	RFnQWD-ESoi0cjEsGNB5mw	64	0	0	0	14.5kb	14.5kb
green	open	test_build_check_20180908	J3g52p_rTqWN09vJtRGfbQ	128	0	241898058	0	104.7gb	104.7gb
green	open	test_build_vc_20180908	yJuvvBBuTpSWn0p8ljwa4A	36	0	1285457	0	553.3mb	553.3mb
green	open	test_build_dataproxy_jx_20180910	sD7-eGlVRd65gAQexwQflQ	128	0	0	0	29.1kb	29.1kb
green	open	test_build_bp_20180908	filwDWWXS9CV6shoYJCxDA	96	0	268364830	0	100.3gb	100.3gb
green	open	test_build_dataproxy_jx_20180907	58pURup2R5S5FFOH6ZJqzg	128	0	454605016	0	241.7gb	241.7gb
green	open	test_build_parser_20180907	uG8ic6xmR4GrxokA5Af5rg	128	0	1718336128	0	823.1gb	823.1gb
green	open	test_build_parser_20180908	7VWslk20QESTKWu-u0mJgA	128	0	552820911	0	258.4gb	258.4gb
green	open	test_build_bp_20180910	P3b7_QW46MS9MQ-UrxzEbQ	96	0	0	0	21.8kb	21.8kb
green	open	test_build_dgw_20180909	7B4iX4kgTLK3MWFsykV2uA	64	0	0	0	14.5kb	14.5kb
green	open	test_build_dgw_20180910	auCvJgCnQmy6_LKUiTWGBg	64	0	0	0	14.5kb	14.5kb
green	open	.kibana	hfnUsiInQDWNuBsBG7RnQA	1	1	52	3	306.2kb	153.1kb
green	open	test_build_check_20180907	pdbxzIenTIq0F8moXD4X_w	128	0	827972452	0	346.2gb	346.2gb
green	open	test_build_mip_20180908	i5kGGFwFSVGfW6OL4_4dOg	64	0	162125	0	46.4mb	46.4mb
green	open	test_build_parser_20180910	ZtA-ERHURH-DsWs_mFO-5A	128	0	0	0	29.1kb	29.1kb
green	open	test_build_check_20180906	35rSJki1Sq074Xn_HG6gKA	128	0	373058345	0	144.7gb	144.7gb
green	open	test_build_vc_20180907	8t8Qab-EQUKpVz43G_04eQ	36	0	2924768	0	1.2gb	1.2gb

阿拉丁建库日志查询

loc:	<input type="text"/>	
siteid:	<input type="text"/>	例如：2010600, 643629 这样的siteid
version:	<input type="text"/>	siteid有多个版本，通过version缩小查询范围, 例如:27123
logid:	<input type="text"/>	check到parser日志有个唯一的logid, 可以通过这个来定位
srcid:	<input type="text"/>	建库资源id, 例如：8355
kv key:	<input type="text"/>	例如：天安门、海底捞
原图URL:	<input type="text"/>	充图原图URL(不加http前缀)
GIPS图URL:	<input type="text"/>	充图后图片地址
增加包/删除包:	<div><div></div></div>	
日期:	<div>20180908</div> 默认是今天，可查询最近3天	
模块:	<div>所有模块</div>	

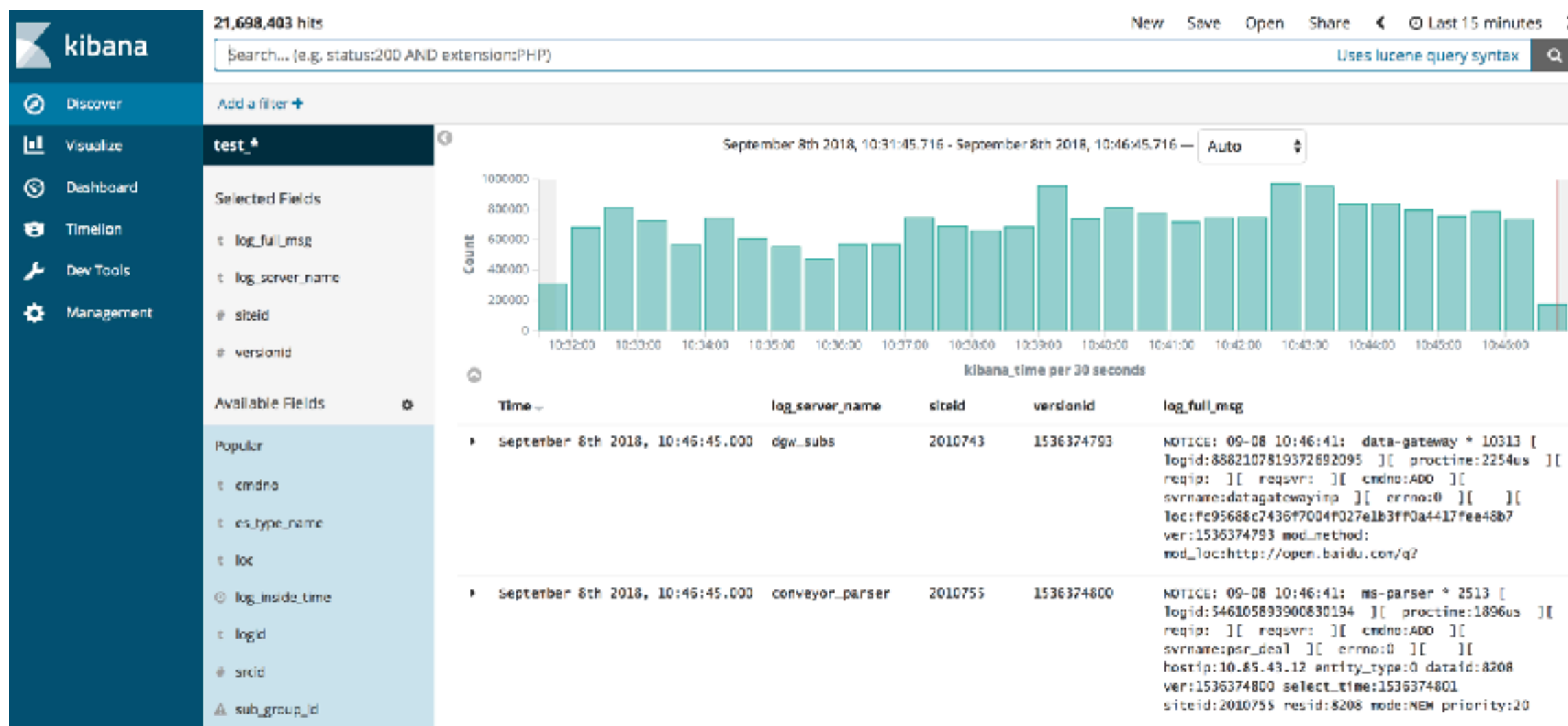
查找

清空

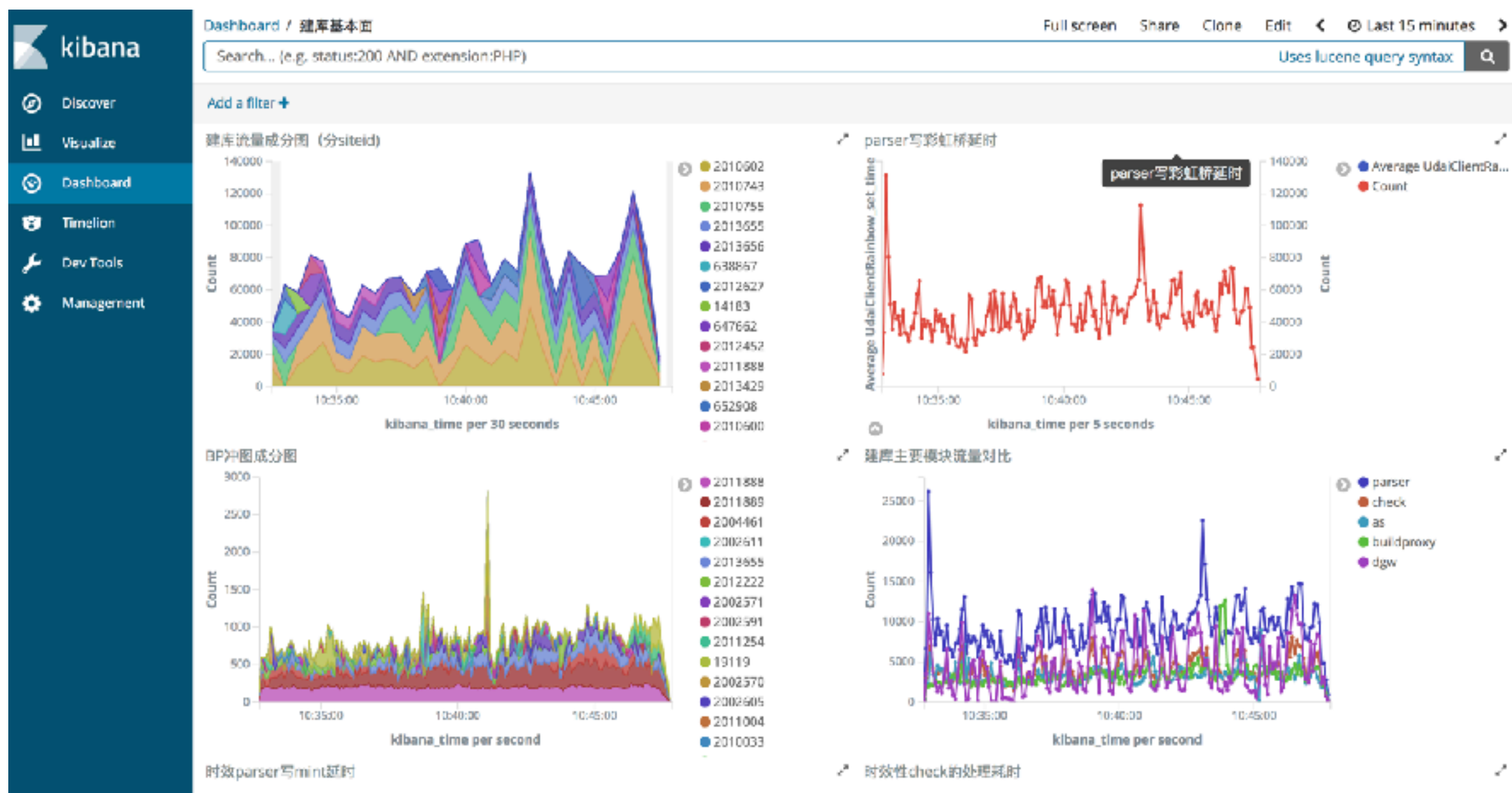
高级查询

Hi群：1595701

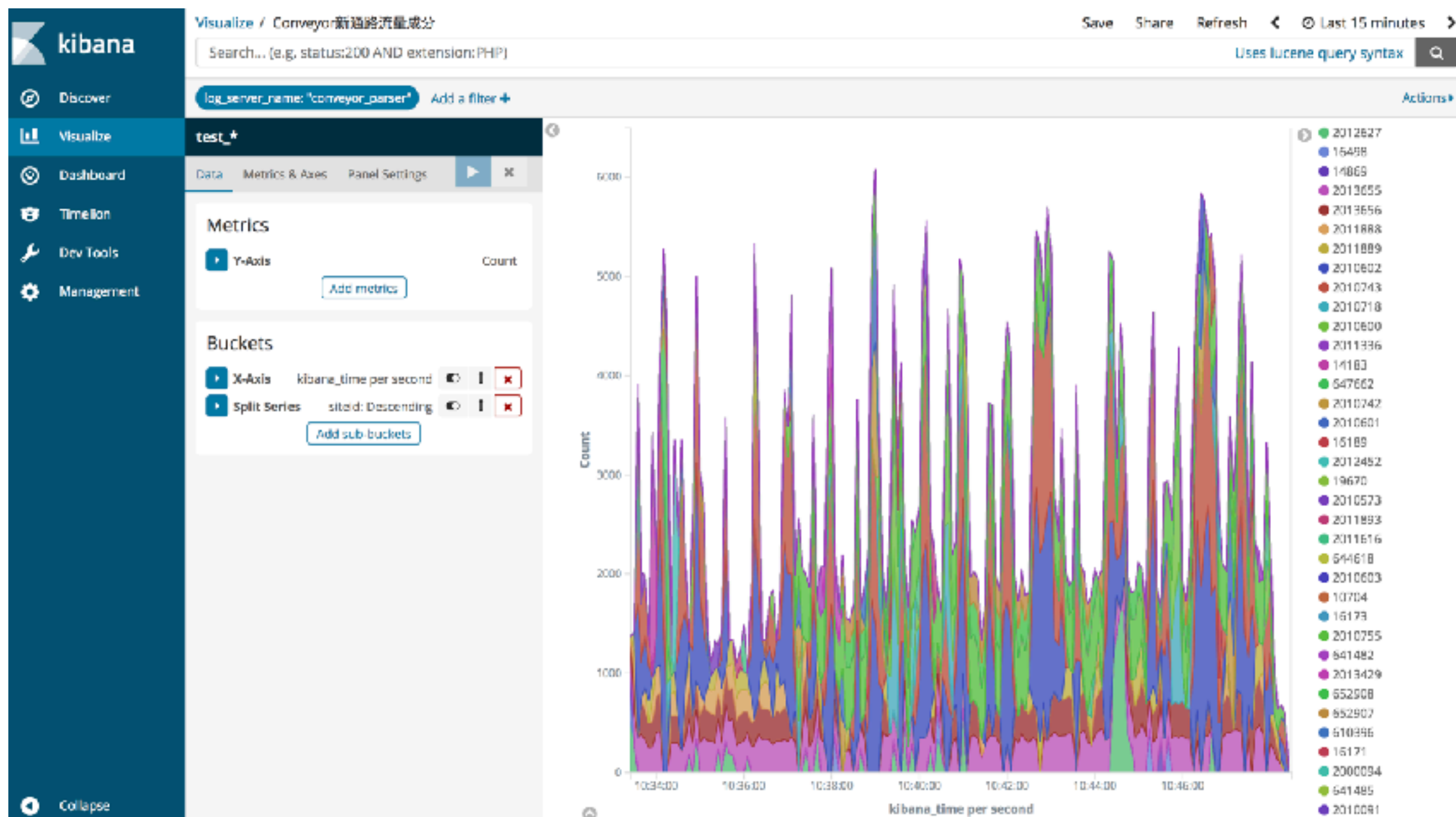
效果



效果



效果



效果

ES集群： 4台物理机（读写接收入口）+ 26个容器，SSD存储。

写峰值： 30Wqps，天级存储日志量25亿条，单副本占存储3TB。

日志保存三天，每天新建之后两天的index，删除三天之前的index。

ES评价

优点

ES集群管理简单，使用Gossip轻松扩容缩容，不需要引入ZK，这点很赞！

ES查询时效性：确实快，case精确查询确实快！

ES shard迁移：shard迁移比较快。当某个容器实例被误删除，单副本的集群，很快大量index都变成red，只能读，不能写入，只能忍痛，执行迁移，降异常分片迁移到正常的es实例上。

缺点： 暂时还没有

期待功能

index shard自动分裂，比如之前设置的shards数比较小，shard所在的ssd存储容量将被打满，可否自动分裂成多shard，到集群的其余实例上。

谢谢



elastic
中文社区

专业、垂直、纯粹的 Elastic 开源技术交流社区
<https://elasticsearch.cn/>