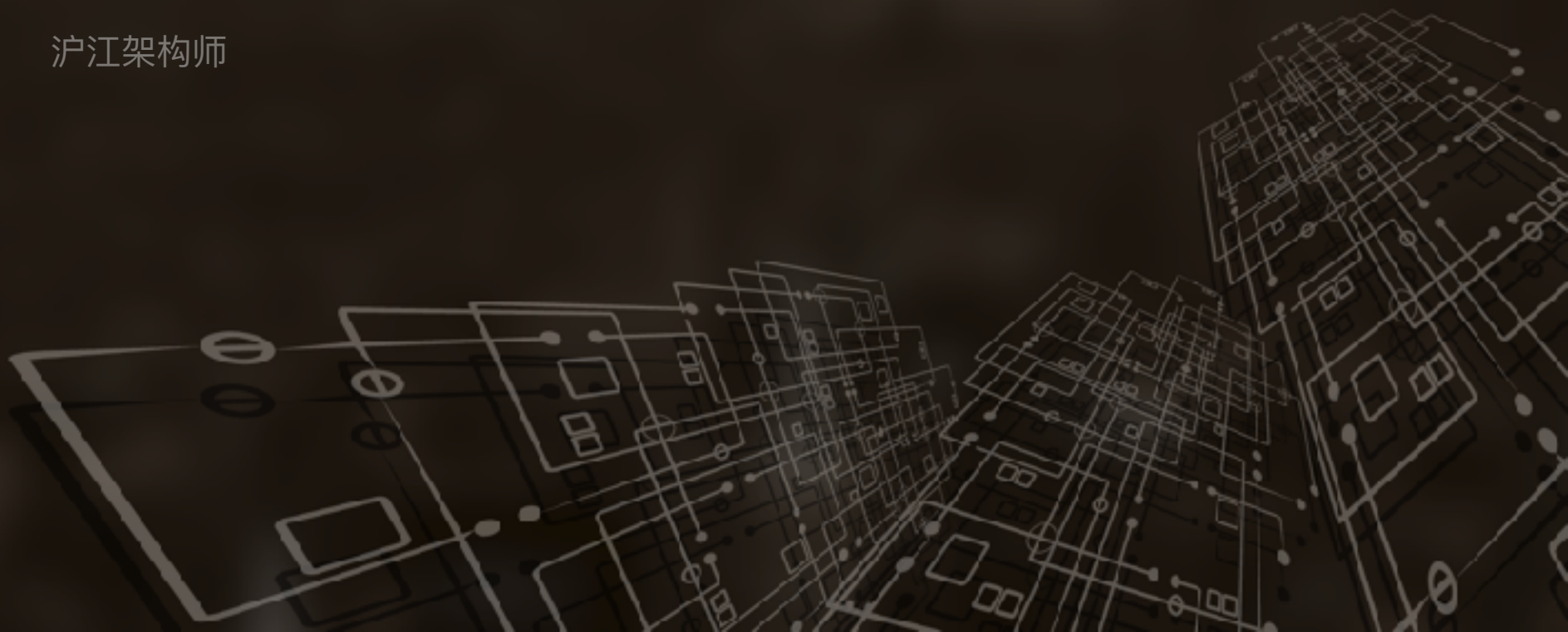


基于 Elasticsearch 的搜索平台

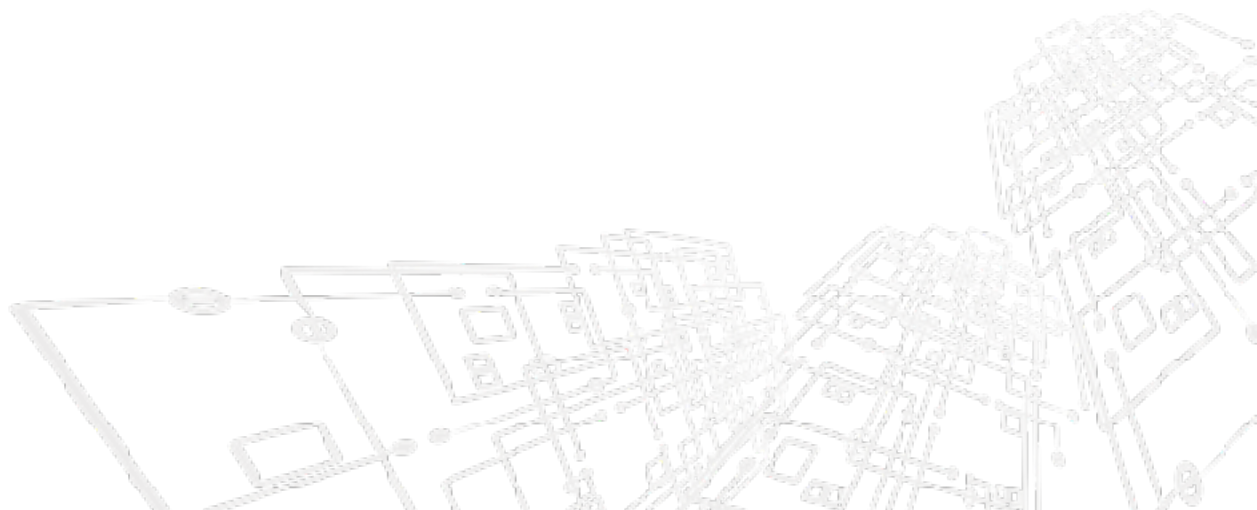
曹林华

沪江架构师

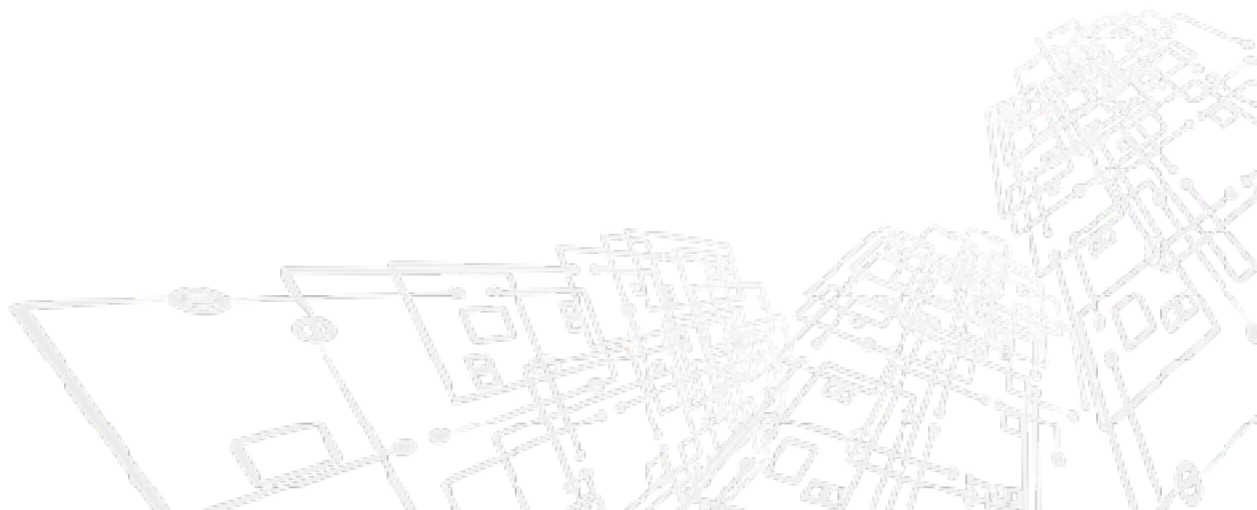


| 个人介绍

- 前百度 高级开发工程师
- 2016年加入沪江，多年技术研发和系统架构经验
- 目前负责沪江搜索平台、日志平台、实验平台、分布式跟踪系统

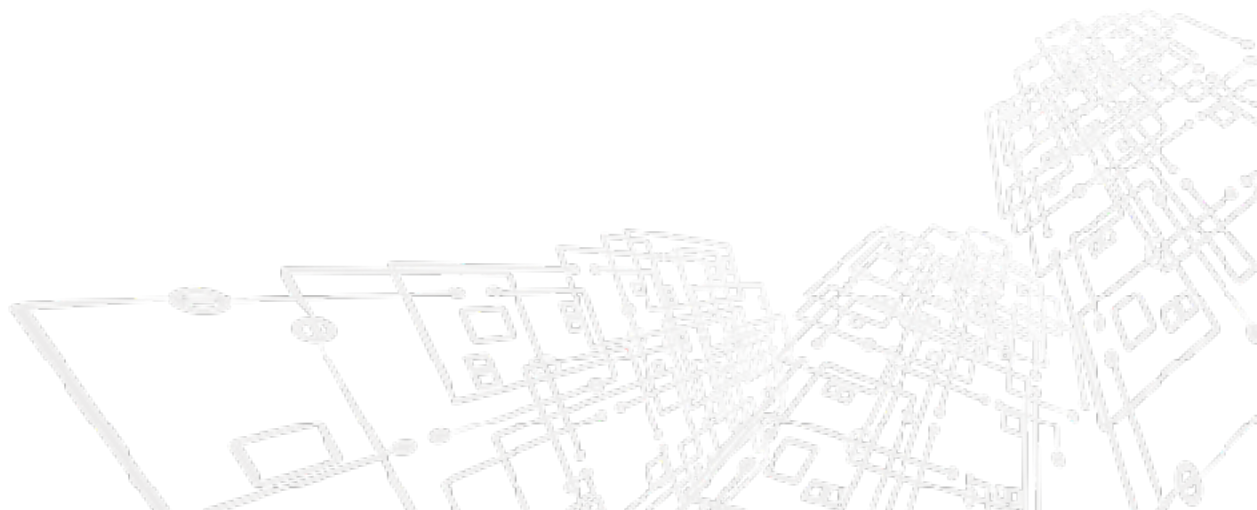


搜索概况



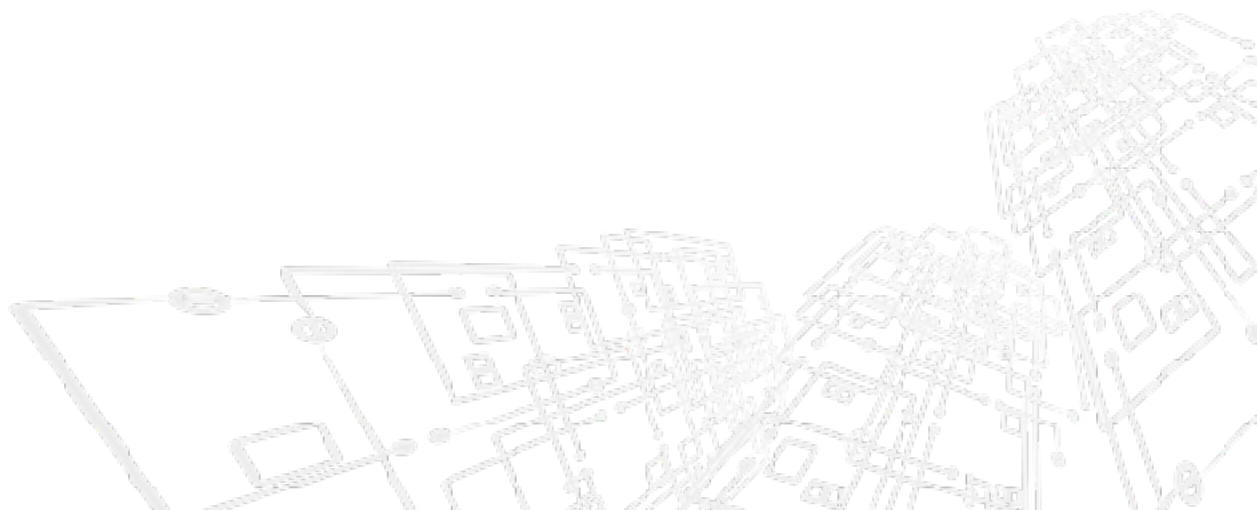
2016年开始，1年时间接入7个业务

支撑了千万级别搜索、亿级别索引，平均耗时25ms

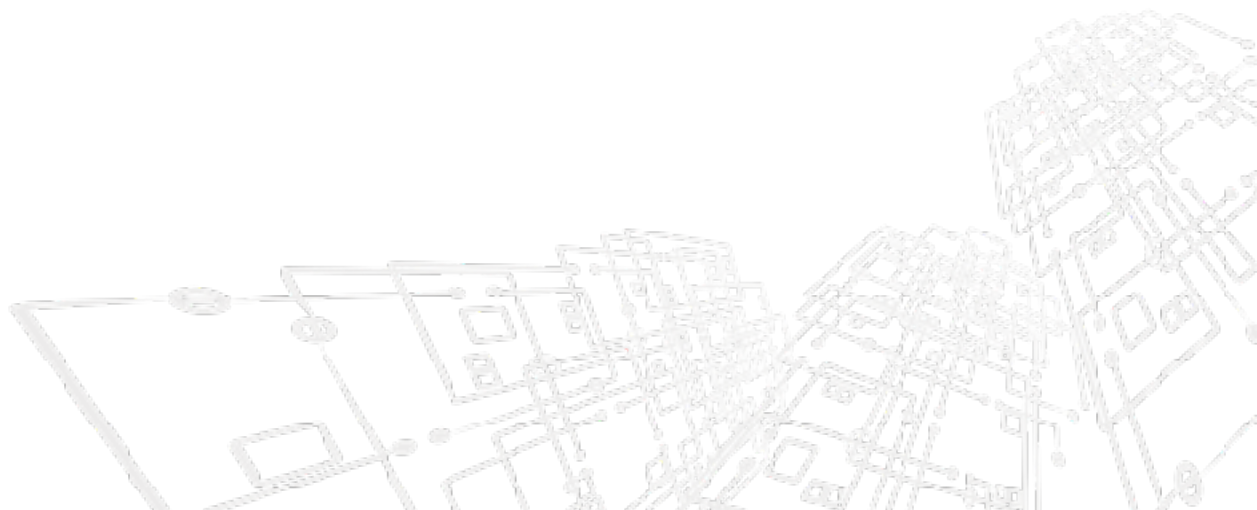


| 目录

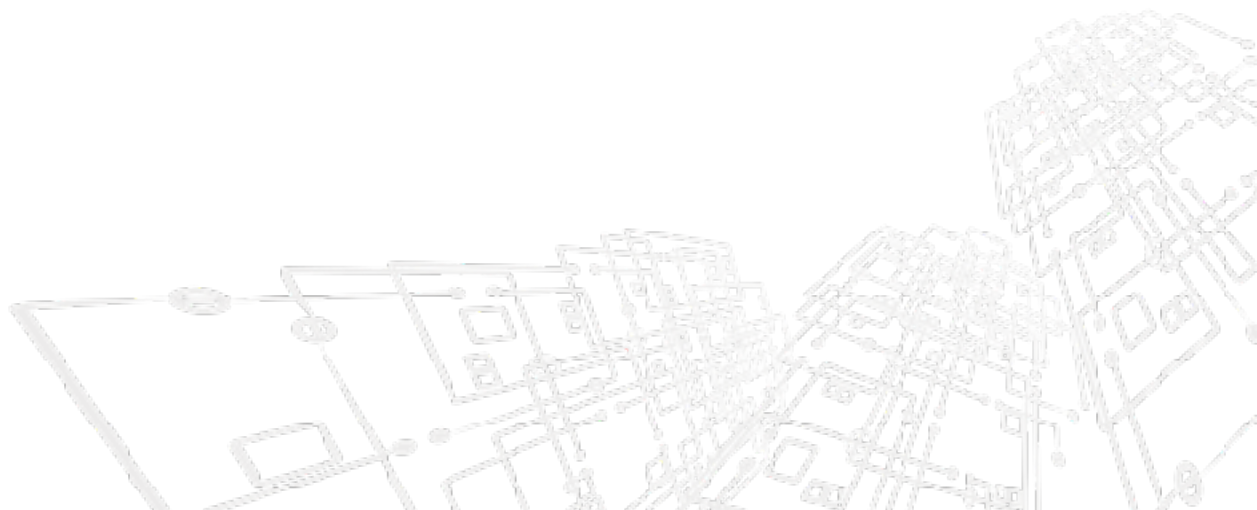
- 产品介绍
- 搜索实践
- 监控保障
- 总结



产品介绍



文章搜索





搜索 查词

英语



搜索结果

Q 搜索相关课程

英语演讲---关于学习英语

学习英语不只是为了增长自己的见识，还为了增加自己的能力，而随着英语的重要性越发明显，这种能力就显得更重要了。下面就是有关学习英语的英语演讲范文，Hi, everyone! My name is XXX. Today my topic is: "I Love English"....

2017-08-20

英语演讲主题

英式英语和美式英语的区别

爱尔兰著名作家萧伯纳说过，英国和美国是被同一种语言分开的两个国家。在漫长的历史演变中，产生了英语的两大分支，即英式英语和美式英语。不管是在元音辅音，还是在重读音节上，美式英语和英式英语都存在差异。美式英语声...

2016-11-14

零基础学习方法 零基础英语课 零基础英语课 美式英语 美式英语

英语常用句型：There be句型

There be句型是一种应用十分广泛和频繁句型。但是对there be句型的多变的特点及其特殊结构的复杂性，并不是每个考生都熟悉了解。在大学英语四级考试中也常常遇到这一句型结构的试题。下面和下边一起来具体看一下吧！ 1、...

2017-08-09

英语常用句型

英语简历范文•英语教师

老师所教科目不同，简历必都不同，今天来看看英语教师的简历吧！Vine No.67, Lane123, Hope Road Job District, Shanghai, China (+86) 138-xxxx-xxxx service@hujiang.com Political Background: probationary party member 敬...

2016-09-23

英语简历模板 教师

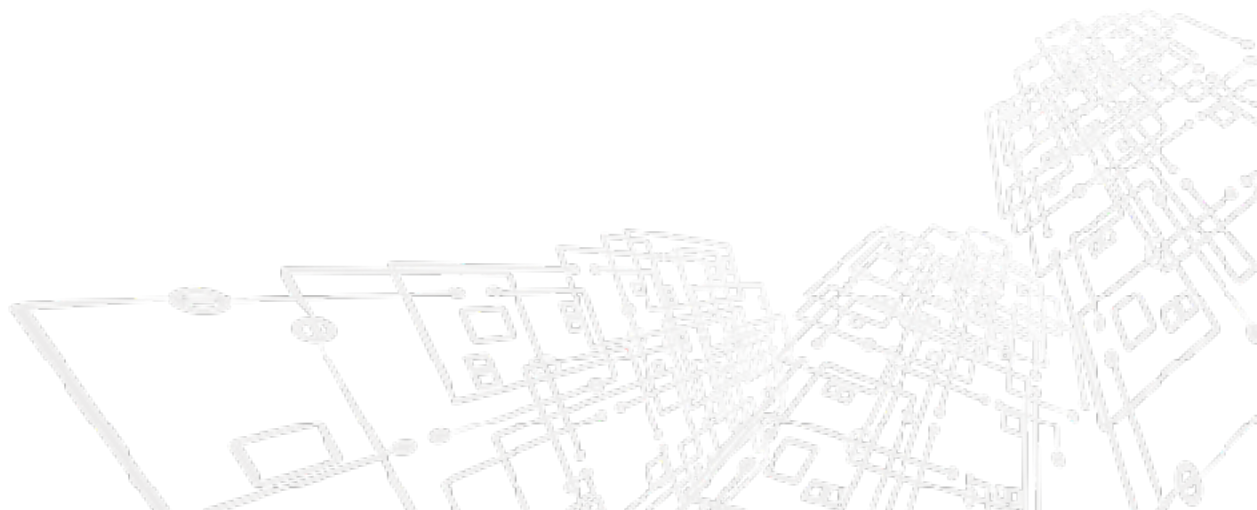
阅读排行榜

- ① 2017年12月英语六级作文模板...
- ② 在中国职场，哪些外语证书含金...
- ③ 有声双语美文：我的生活，充满...
- ④ 有声双语美文：善念，很可能是...
- ⑤ 有声双语美文：智慧无处不在，...
- ⑥ Quora精选：一辈子单身很难吗？
- ⑦ 有声双语美文：即使囊中羞涩，...
- ⑧ 经验：四级688分，点滴积累才...

热搜榜

- | | |
|-----------|--------|
| ① 翻译英文怎么说 | 330018 |
| ② 小说英文怎么说 | 195567 |
| ③ 电影英文怎么说 | 122063 |
| ④ 音乐英文怎么说 | 69437 |
| ⑤ 新闻英文怎么说 | 59378 |
| ⑥ 苹果英文怎么说 | 58884 |

用户、群搜索





用户

全部6741个用户 >



沪江留学

用户名: 沪江留学

188

关注

644

粉丝

21

节目



沪江德语

用户名: 沪江德语

24

关注

516

粉丝

17

节目



沪江

用户名: 沪江网

3

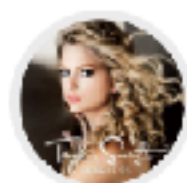
关注

13

粉丝

0

节目



沪江妮了

用户名: 沪江妮子

0

关注

0

粉丝

0

节目

群

全部109456个群 >



沪江法语学习大厅

群号: 233233

5534个成员



沪江西语学习大厅

群号: 233332

3506个成员



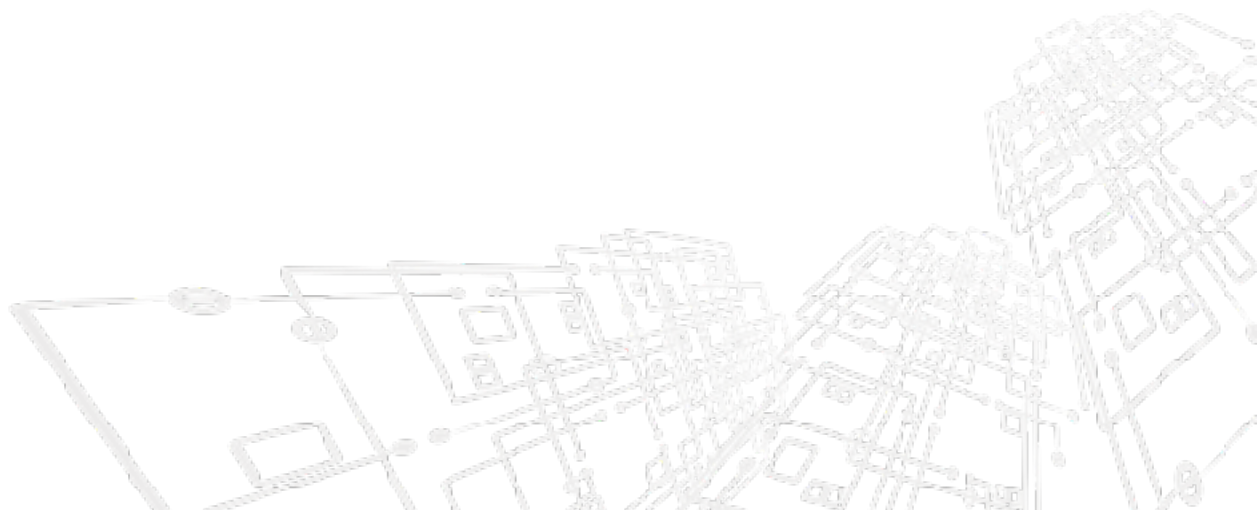
沪江网校法语同学会

群号: 81508153

1656个成员



课程搜索





英语

搜索

★我收藏的课程

全部课程 > “英语”

共 626 个结果

全部 英语 小语种 出国留学 考研 财会金融 职场兴趣 中小幼教育

当前水平: 零基础 低水平 有基础 初中水平 高中水平 大学及以上水平 其它

班型: 1对1 VIP定制班 小班 大班 协议班 体验班 其它班型 精英班 副班 名校班 基础培优班 考前专项班

授课模式: 课件课 直播课 视频课 混合课 直播+课件

开班时间: 报班即学 10月开班 11月开班 12月开班

综合 价格 人气

< 1 / 32 页 >



新概念英语1、2、3、4册连读【学霸班】 ②免费试听

【5.5折最后2天】送教材，配套人工答疑+作业批改服务，直达大学英语六级水平

¥1395.9

随到随学有效期 490 天 | 总366课时 老师: Amy老师, Angela老师, Sue老师, Ada老师, Lillian老师

查看详情



BEC商务英语初、中、高级连读【签约学霸班】 ②免费试听

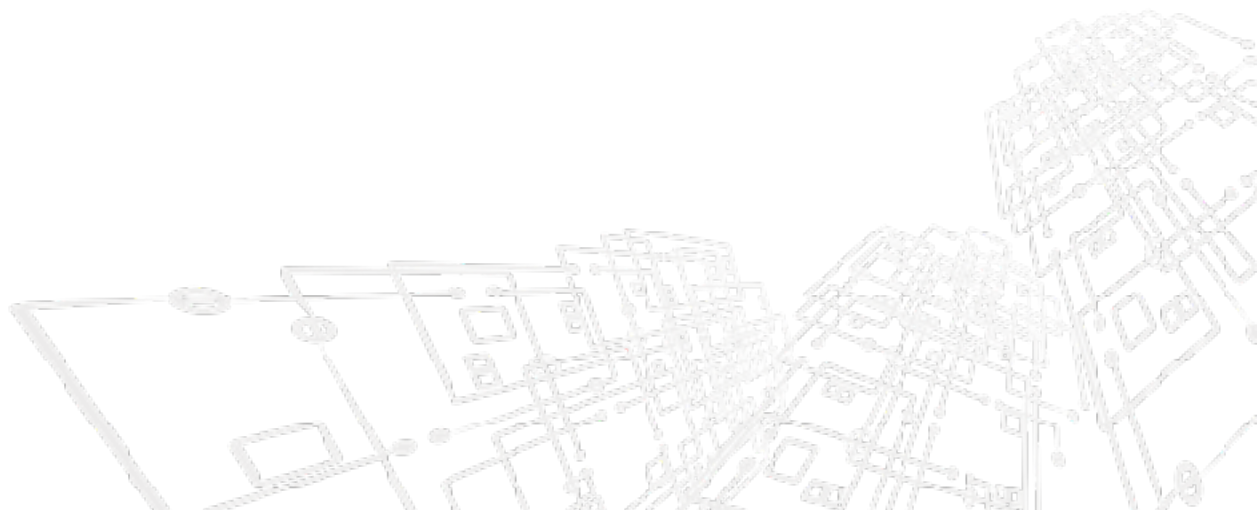
【学霸班7.2折 最后2天】职场实用英语+考证，签协议，不过还能免费重读。

¥2951.3

报名即学至 2018-12-19 | 总288课时 老师: Gabby老师, Dennis老师, Eric老师, 小兜老师, Dolly老师, Anna老师

查看详情

搜索遇到的问题





搜索效果不好

中英日文搜索效果不满意



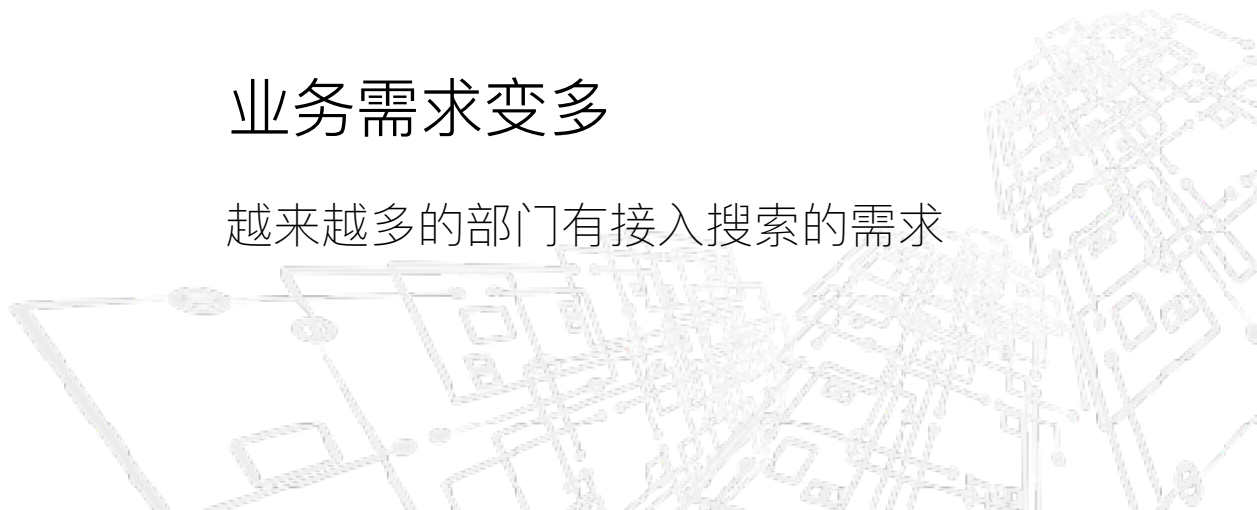
跨库跨表无法查询

数据量变多，需要采取分库分表查询

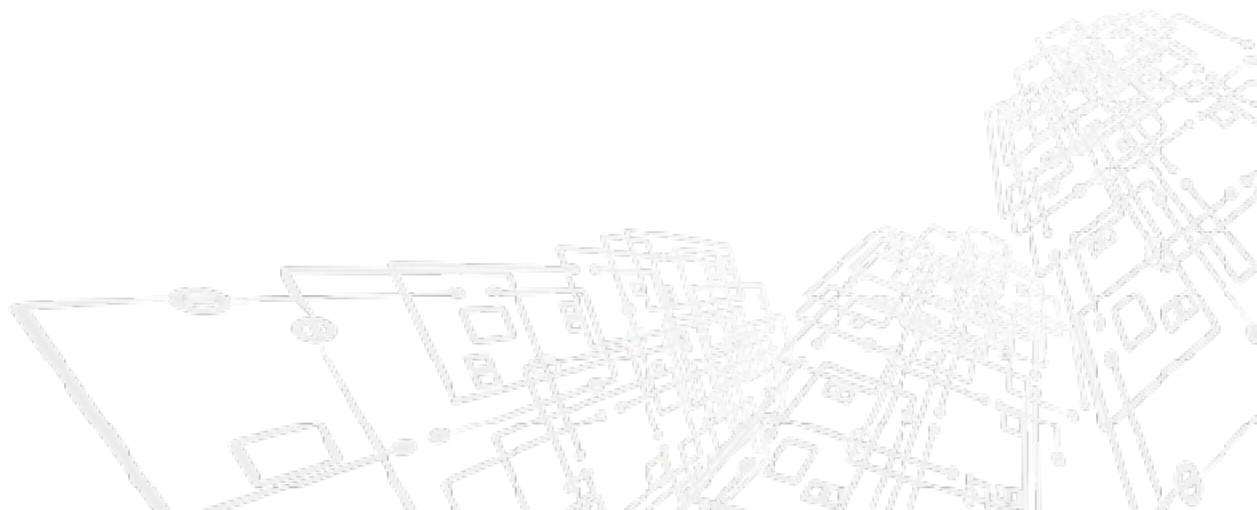


业务需求变多

越来越多的部门有接入搜索的需求



搜索实践

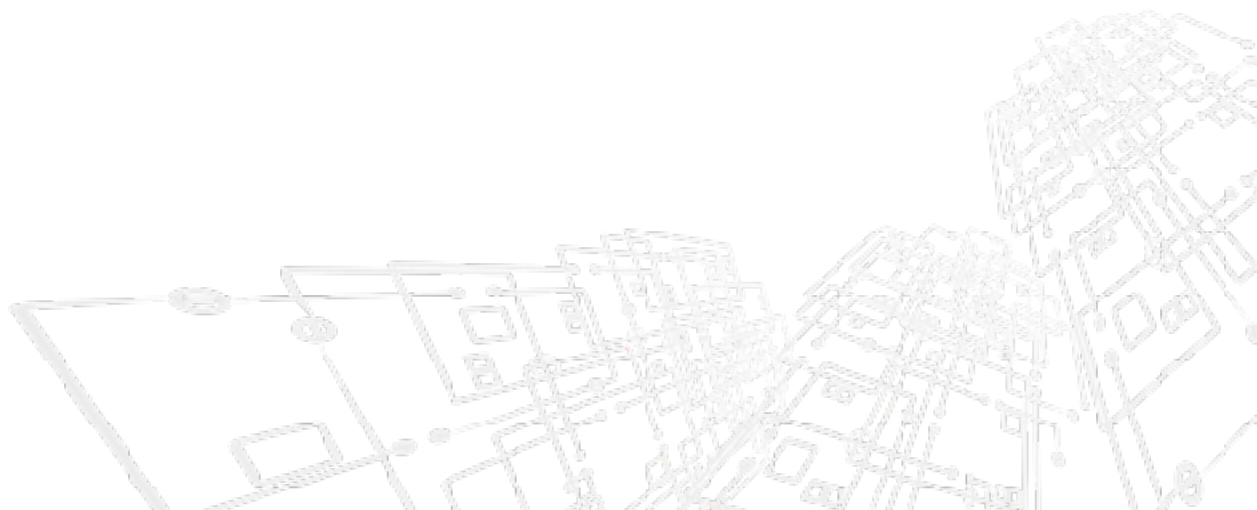




elasticsearch



What





全文搜索



分布式高可用



RESTful API



近实时搜索与更新

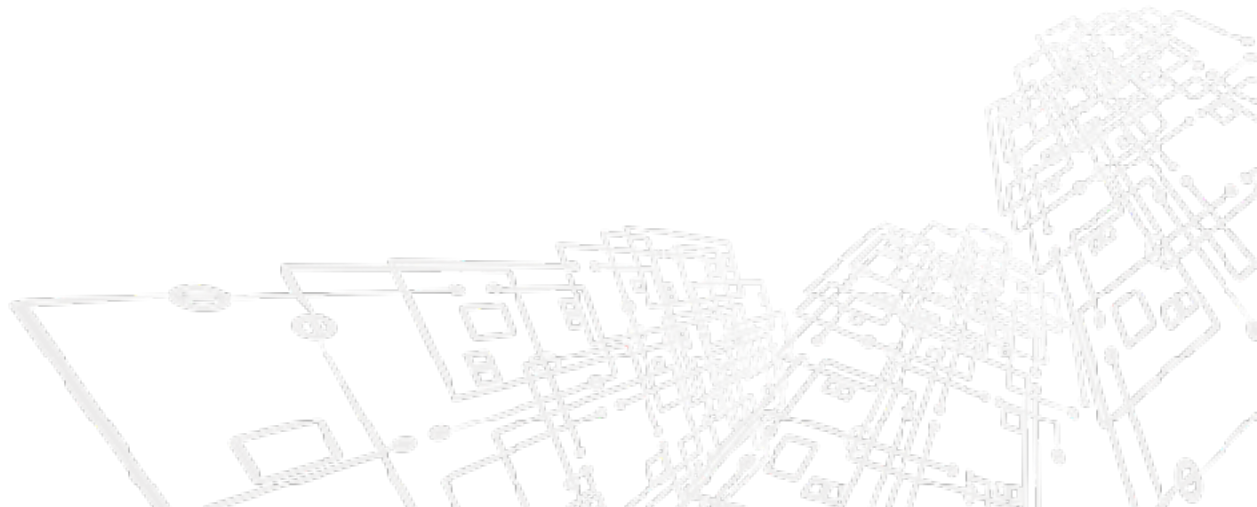


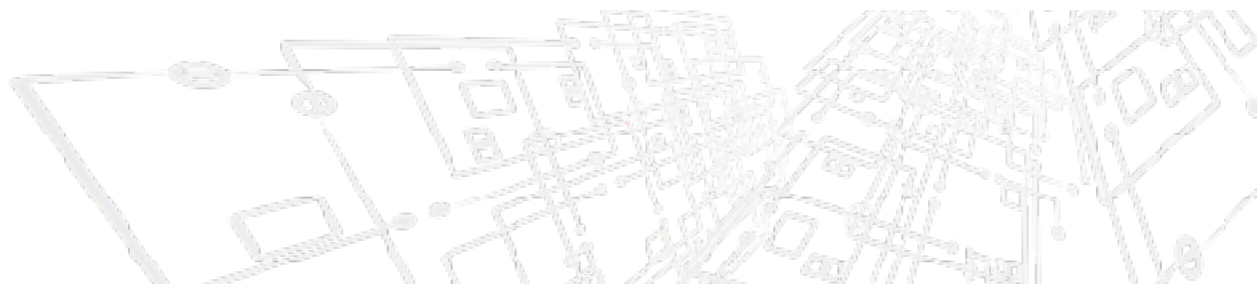
面向文档



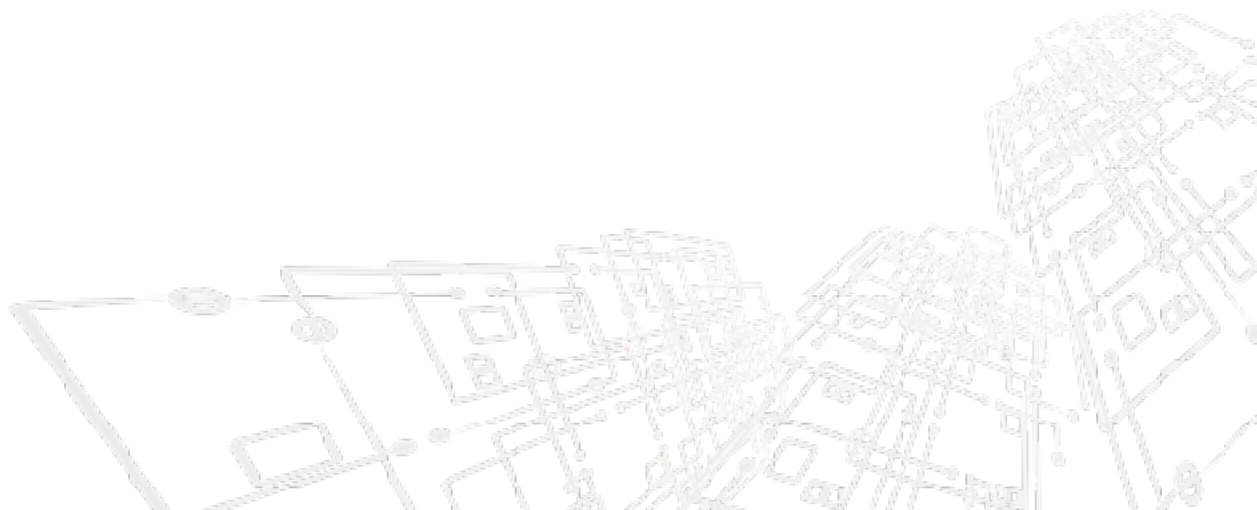
Schema free

Who





整体架构



业务线

主站

网校

IM

CRM

搜索平台

Index

Search

AB Test

Elasticsearch

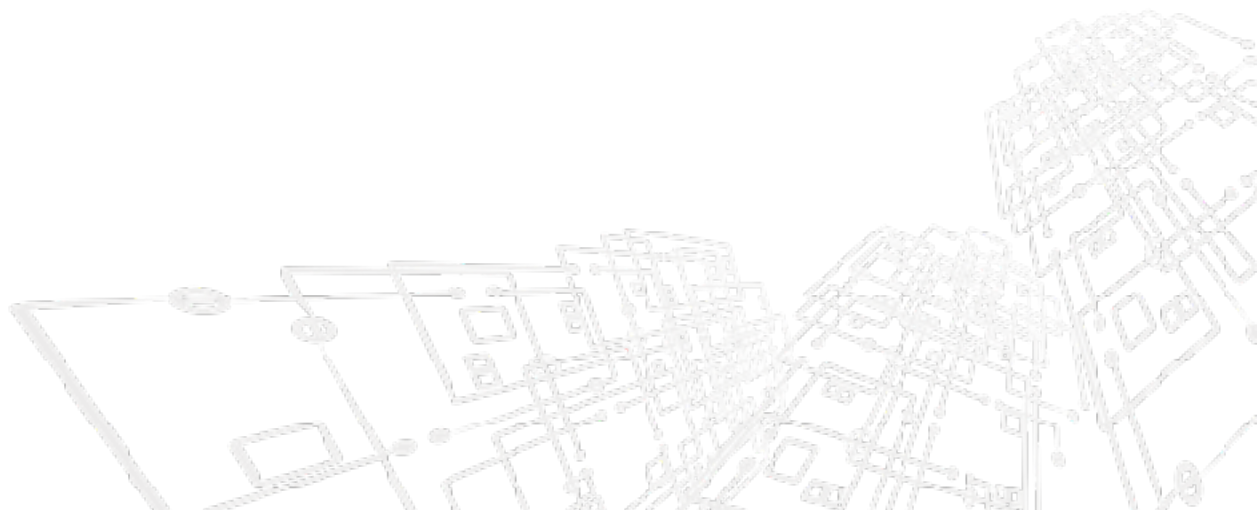
监控

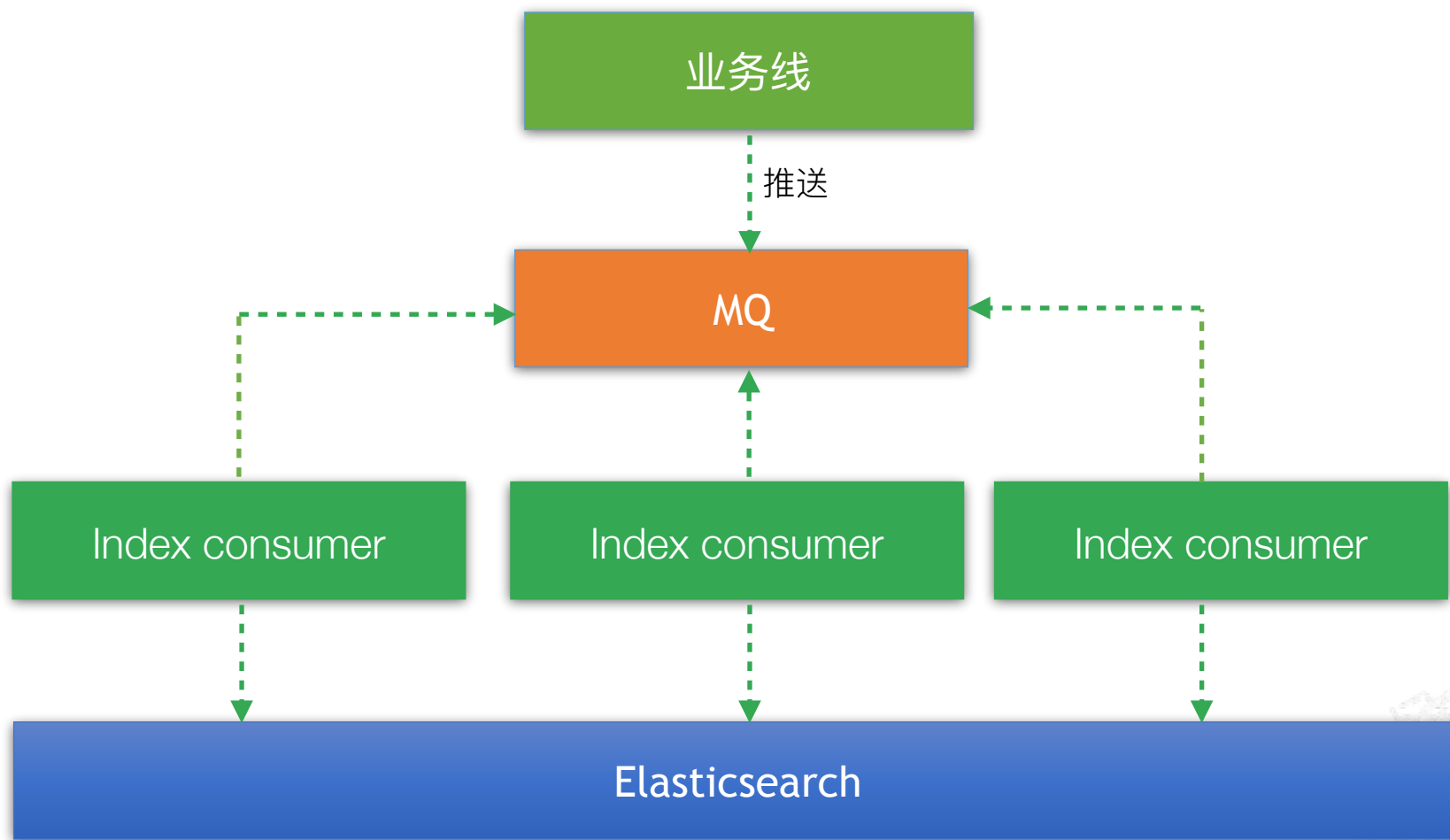
数据层

BI

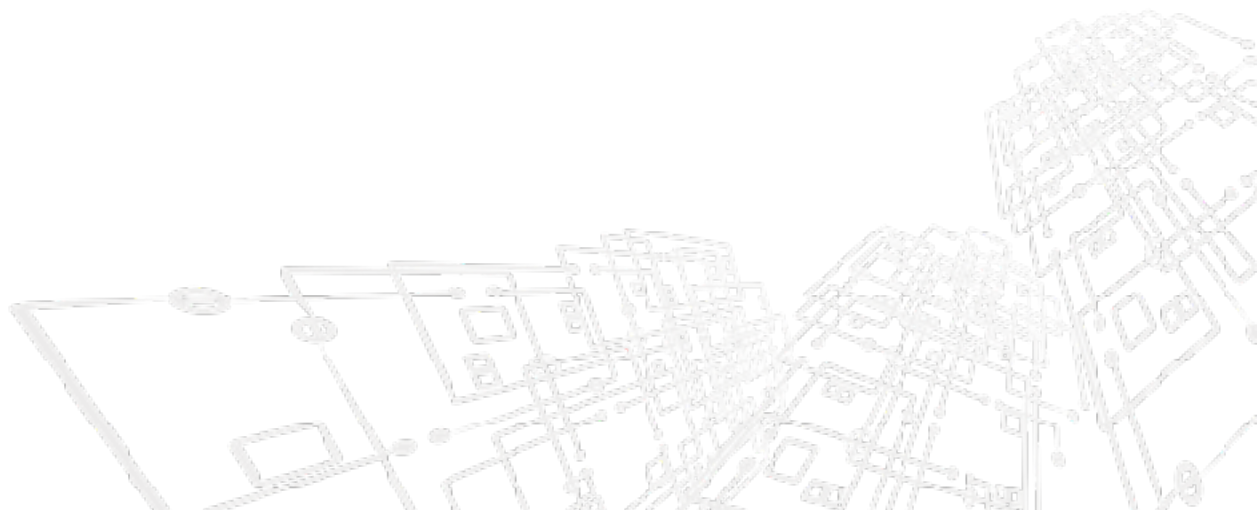


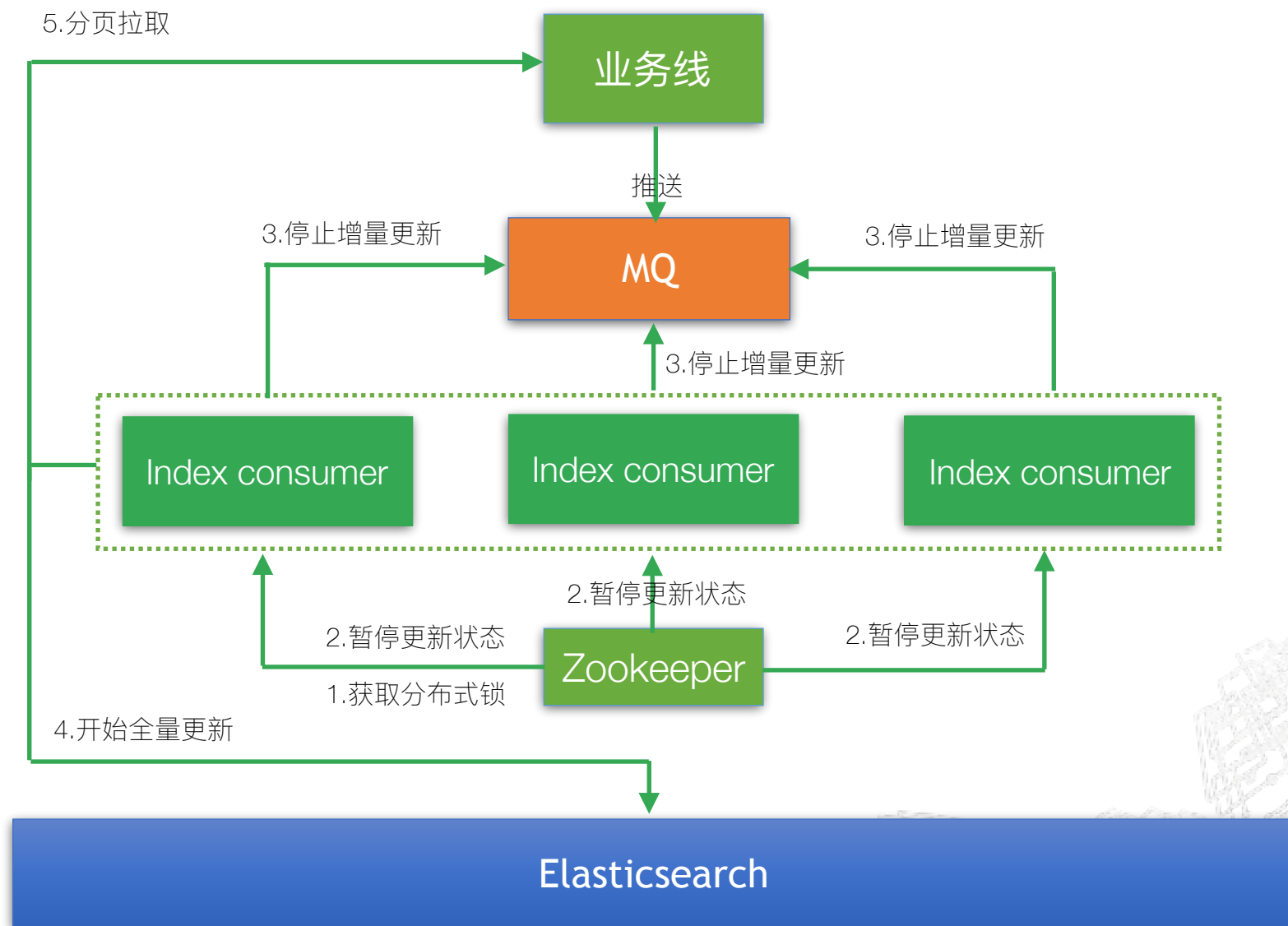
增量更新



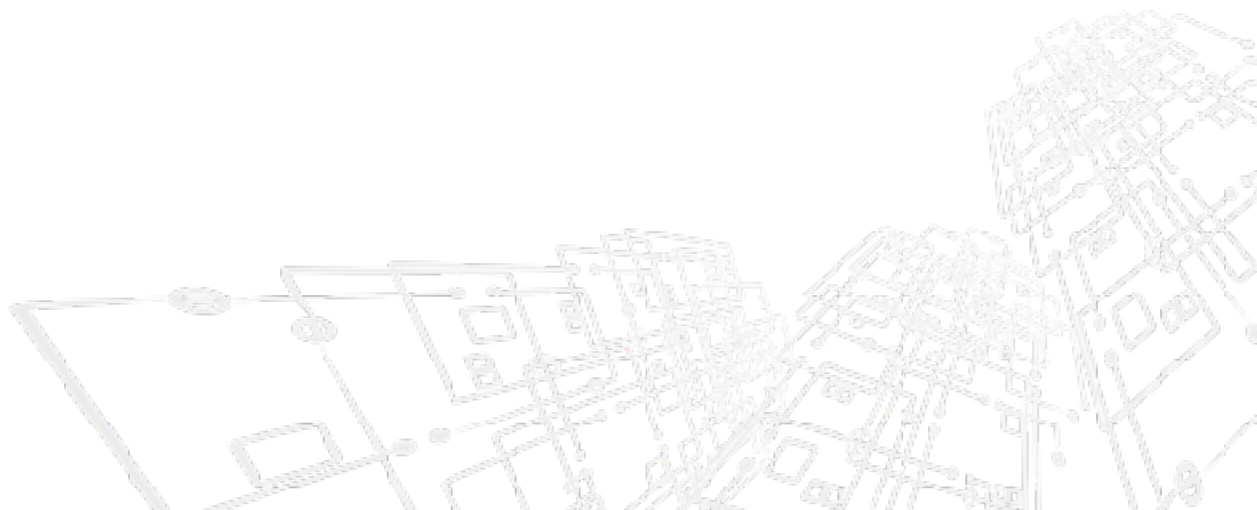


全量更新





集群扩容



扩容前

NODE 1

NODE 2

NODE 3

NODE 4

NODE 5

扩容中

NODE 1

NODE 2

NODE 3

NODE 4

NODE 5

扩容后

NODE 1

NODE 2

NODE 3

NODE 4

NODE 5

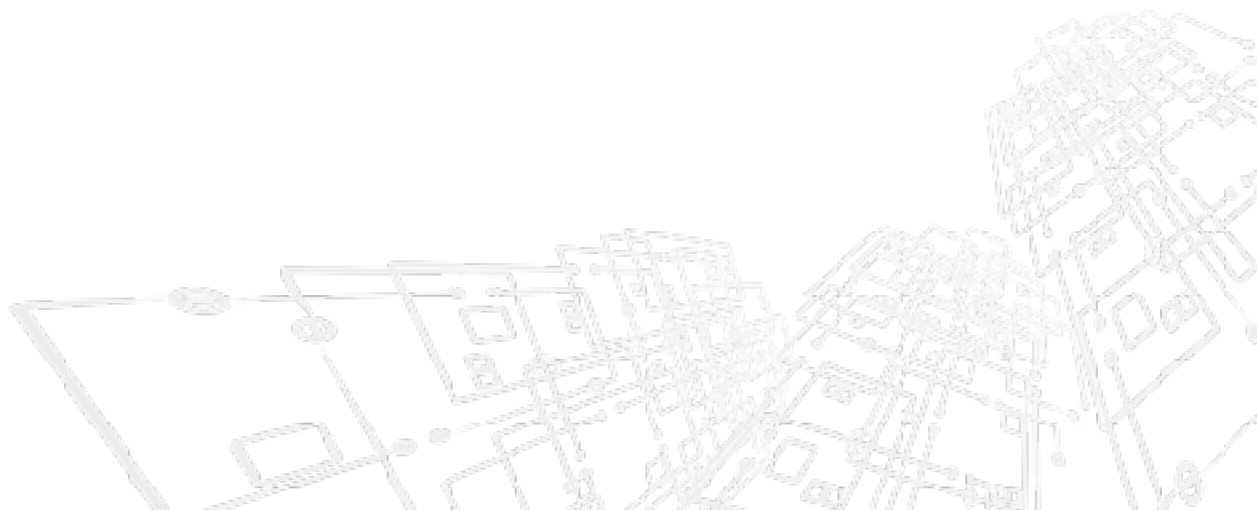
Add Data Node



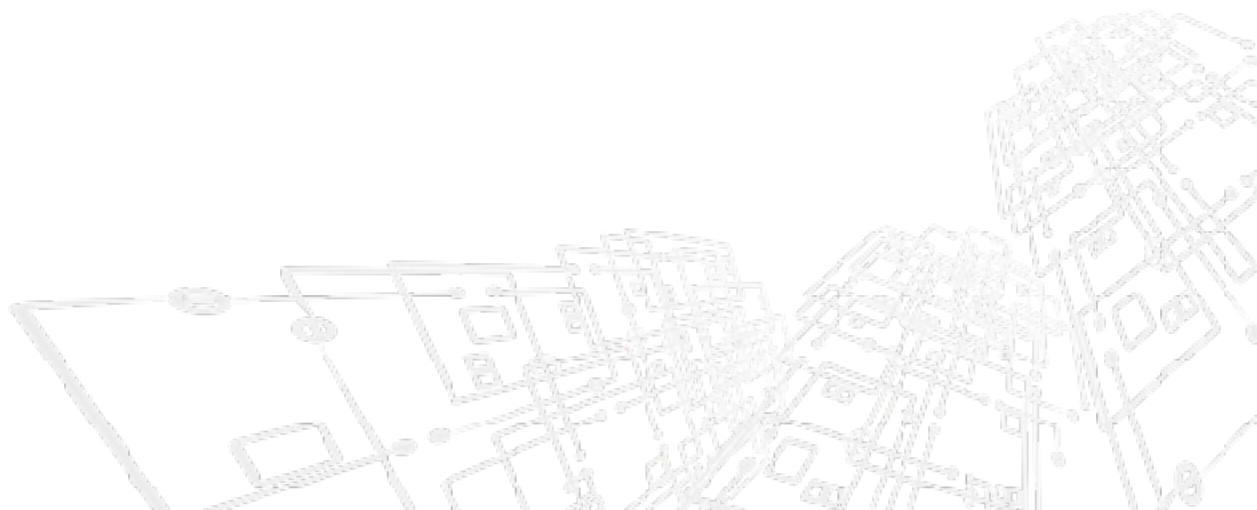
Restart Master Node



监控保障



ZABBIX



Zoom: 5m 15m 30m 1h 2h 3h 6h 12h 1d 3d 7d 14d 1m 3m 6m 1y All

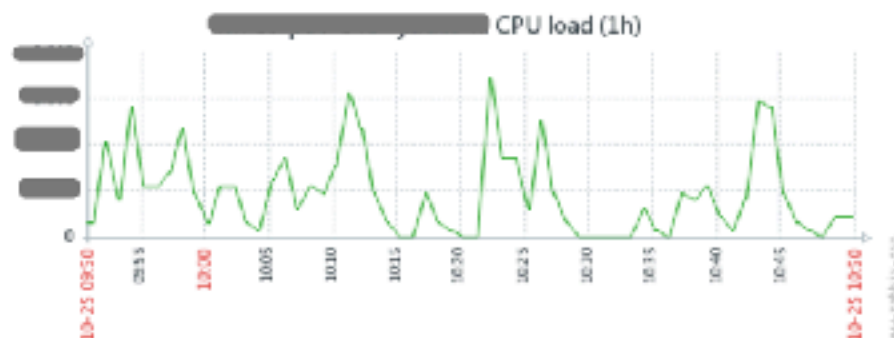
2017-10-25 09:50 - 2017-10-25 10:50 (now)



<< 1y 3m 1m 7d 1d 12h 1h 5m | 5m 1h 12h 1d 7d 1m 6m 1y >>



1h Fixed



■ Processor load (1 min average per core) [avg]
■ Processor load (5 min average per core) [no data]
■ Processor load (15 min average per core) [no data]

last min avg max

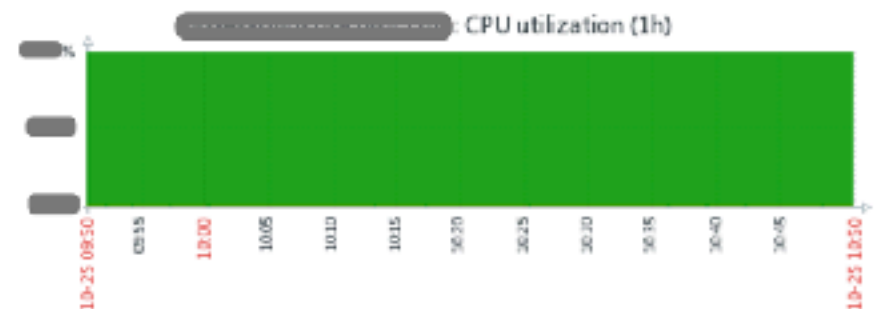
Data from history. Generated in 0.38 sec.



■ tcpv4-connection-established [all]

last min avg max

Data from history. Generated in 0.38 sec.



■ CPU idle time [avg]
■ CPU user time [no data]
■ CPU system time [no data]
■ CPU iowait time [avg]
■ CPU nice time [no data]
■ CPU interrupt time [no data]
■ CPU softirq time [no data]
■ CPU steal time [no data]

last min avg max

Data from history. Generated in 0.12 sec.



■ Available memory [all]

last min avg max

Data from history. Generated in 0.38 sec.

Head

<https://github.com/mobz/elasticsearch-head>



Elasticsearch

返回 筛选器值: green (90 of 90)

信息 -

概览 索引 数据浏览 基本查询 (+) 复合查询 (-)

集群概览 刷新索引 排序索引 查看别名 索引过滤器

刷新 -

portal_suggest_20170829165428
size:
docs:
[信息](#) [动作](#)

portal_20170918110456
size:
docs:
[信息](#) [动作](#)

crm_20170901142454
size:
docs:
[信息](#) [动作](#)

con_question_20170504
size:
docs:
[信息](#) [动作](#)

con_question

portal_new X

portal_suggest X

crm X

● node149	1 2		3 4		3		0		4
● node150		2 3		0 1		0 1		0 1	
● node204	0 1		0		4	0			3 4
★ node231		3 4		2 3		2 3		2 3	
● node99	0		4		1 2		1 2		1 2

Bigdesk

<https://github.com/lukas-vlcek/bigdesk>



HTTP & Transport

HTTP address:

Transport address:

Bound address:

Bound address:

Publish address:

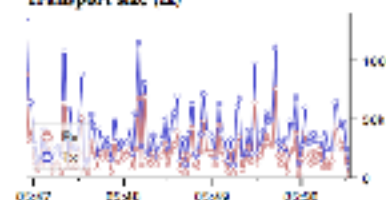
Publish address:

Channels



Transport: 10B
HTTP: 0
HTTP total opened: na

Transport size (Δ)



Series:
Rx:
Tx:

Indices

Docs count:

Flush:

Size:

Docs deleted:

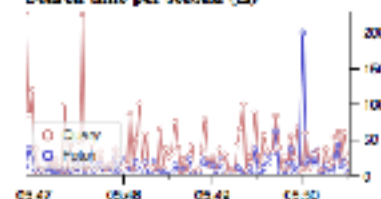
Refresh:

Search requests per second (Δ)



Query:
Fetch:

Search time per second (Δ)



Query: 18h
Fetch: 5.2h

Get requests per second (Δ)



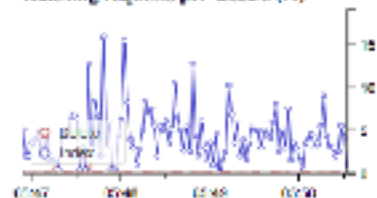
Get: 182
Exists: 54
Missing: 128

Get time per second (Δ)



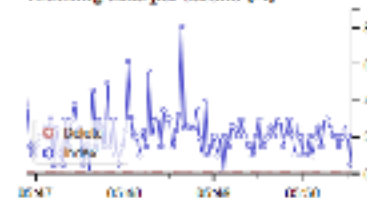
Get: 127ms
Exists: 53ms
Missing: 74ms

Indexing requests per second (Δ)



Delete: 27531

Indexing time per second (Δ)



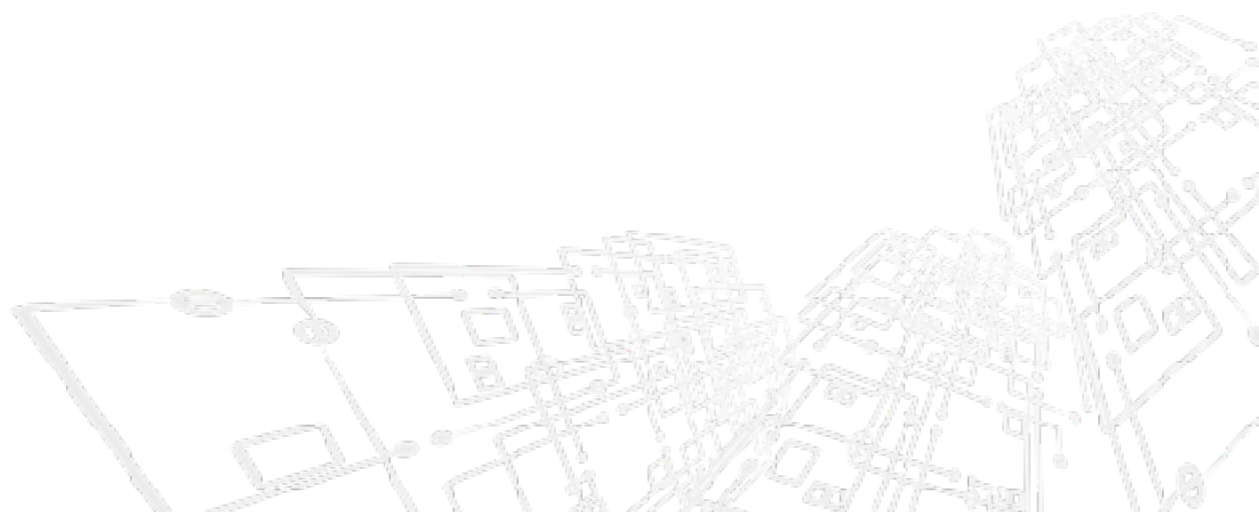
Delete: 3.5s

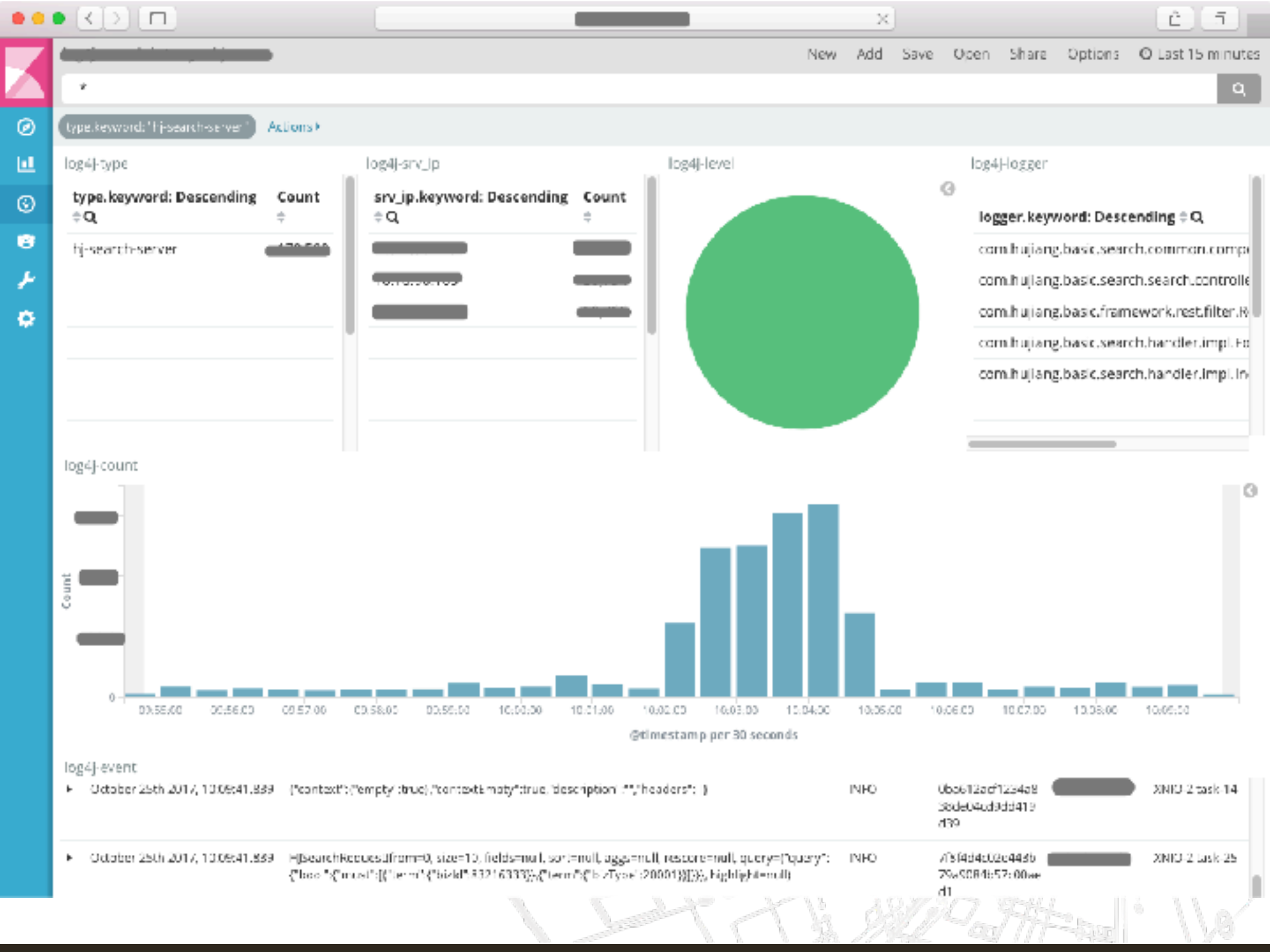
ID: n/a

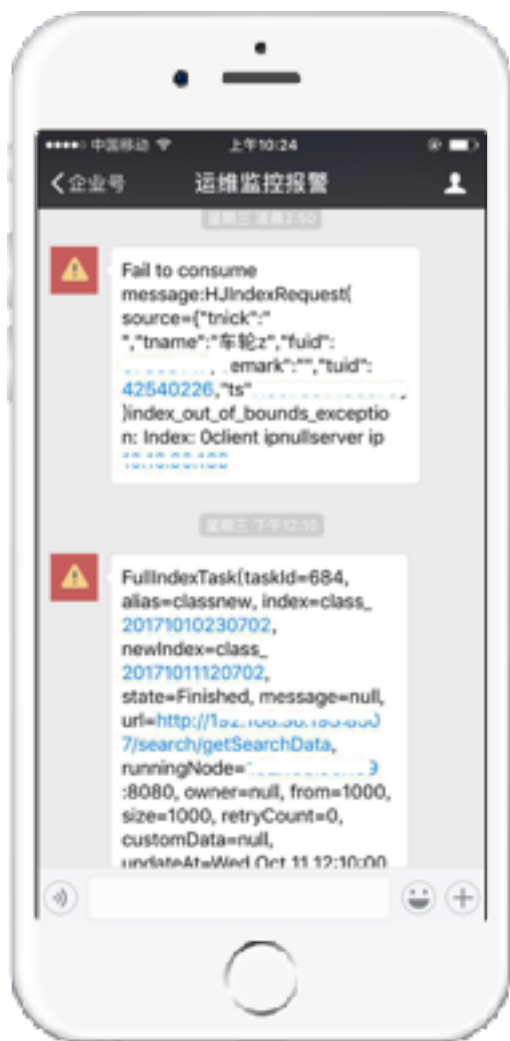
Filter: n/a



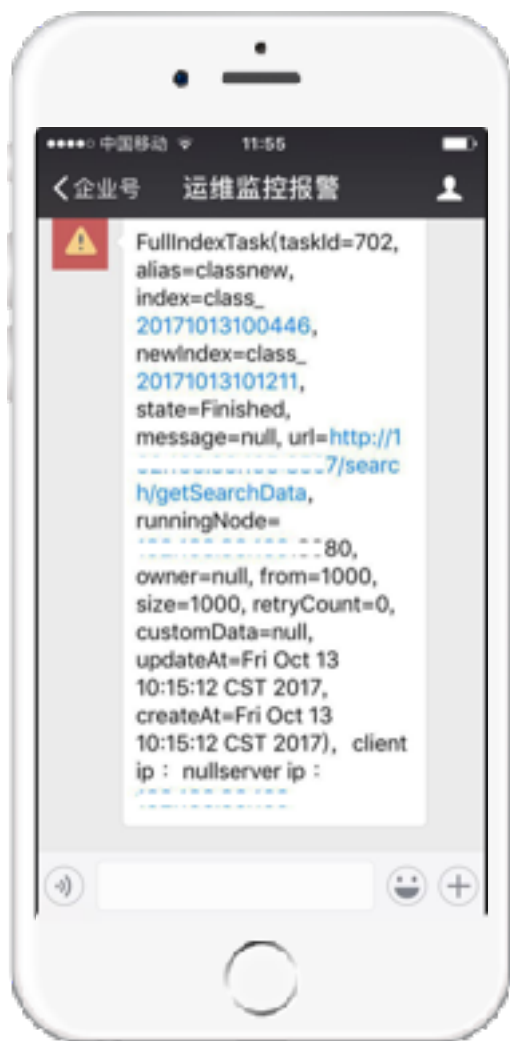
elastic





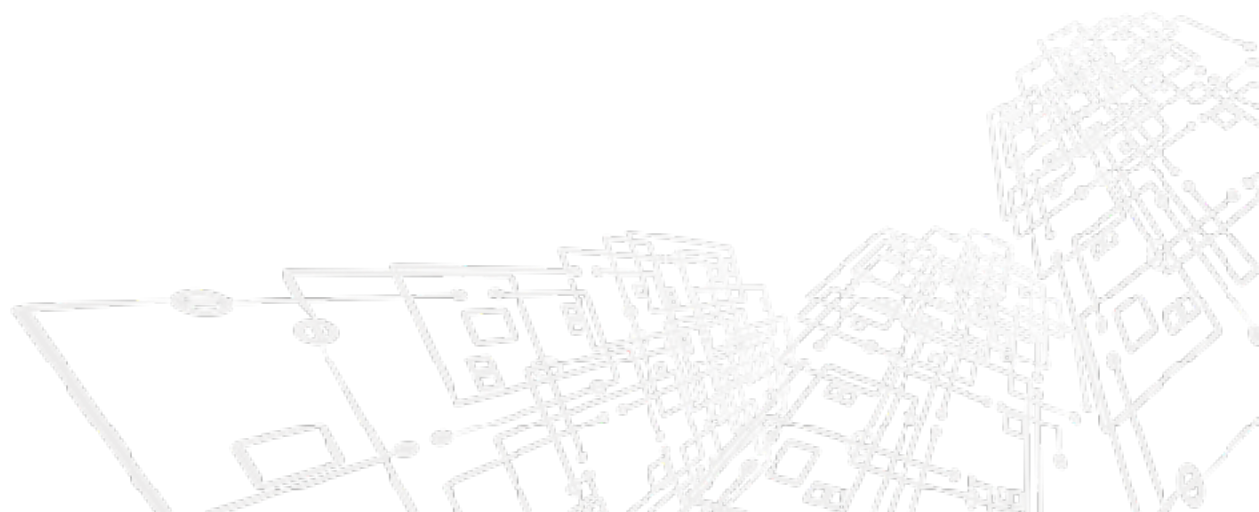


增量更新错误

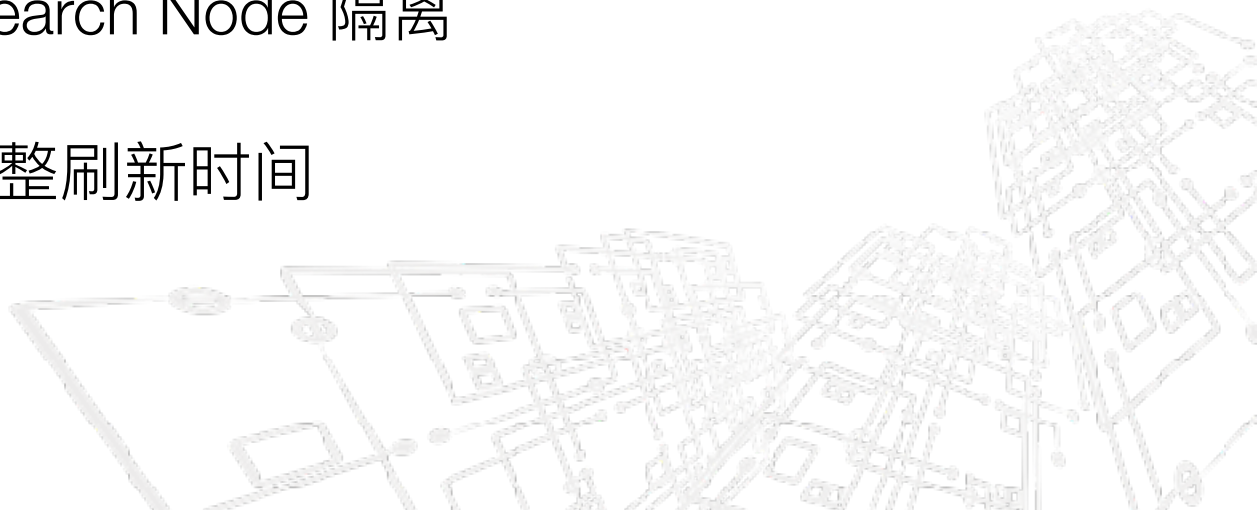


全量更新通知

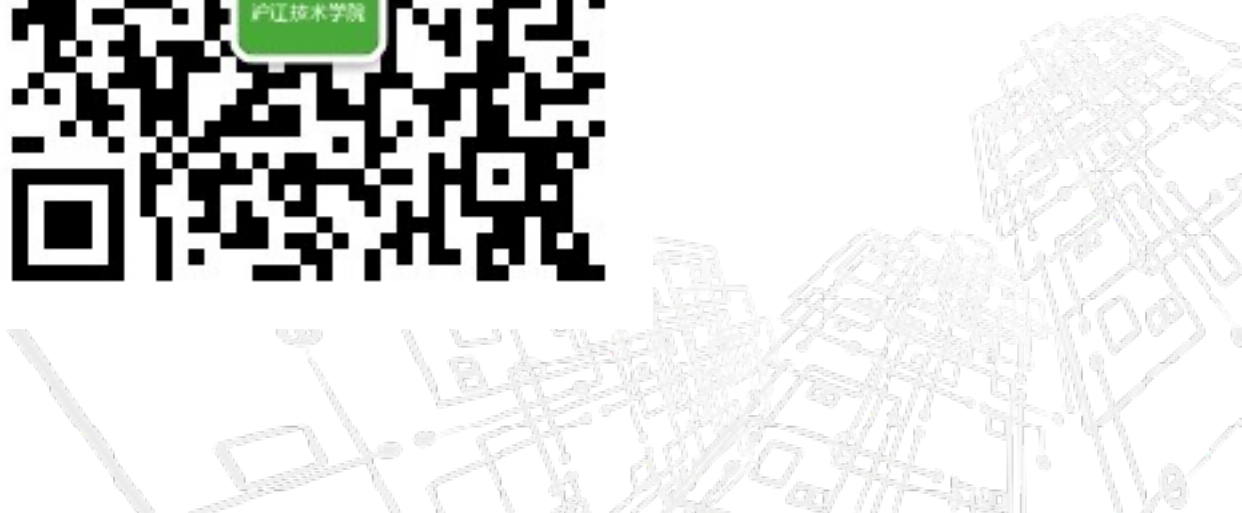
再说两句



- 批量写索引
- 分片数不是越多越好
- 避免使用wildcard的搜索“***ABC**”
- 尽量少用高亮，会导致CPU使用过高
- Index Node 与Search Node 隔离
- 根据数据量来调整刷新时间



Q & A





elastic
中文社区

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