

# Elasticsearch 通用搜索平台

马华标 城破



# 总序

- 什么是通用搜索平台
- 架构演进过程
- **Elasticsearch 怎么玩**



# 平台定位

蚂蚁金服

阿里云

天猫 ...

业务团队（应用、系统、产品）

通用搜索平台（中台）

搜索引擎（基础服务）





# 第一代平台 - ZSearch

```
// 建索引
```

```
curl -XPUT 'http://search.alipay.com:9200/user/1' -d '{  
  "user" : "kimchy",  
  "message_text" : "trying out ZSearch"  
}'
```

```
// 查数据
```

```
curl -XPOST http://search.alipay.com:9200/user/_search -d '{  
  "user": "kimchy"  
}'
```



# 第一代平台 - 引擎移植



elasticsearch

+

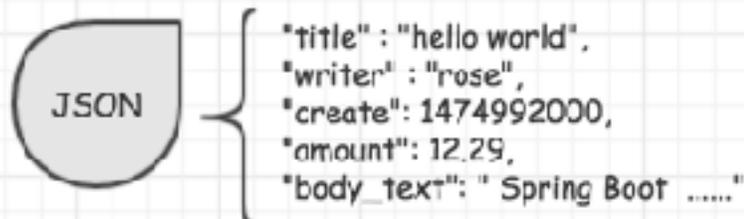


集团搜索引擎HA3

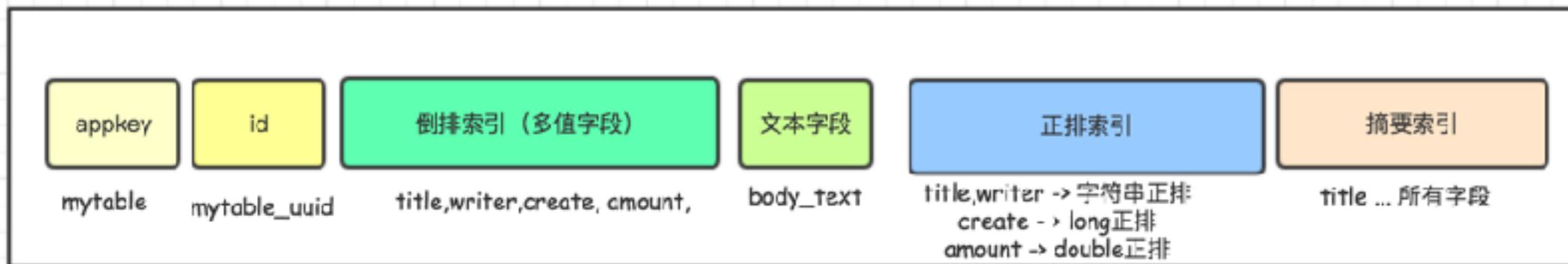


# 自动映射 - 宽表技术

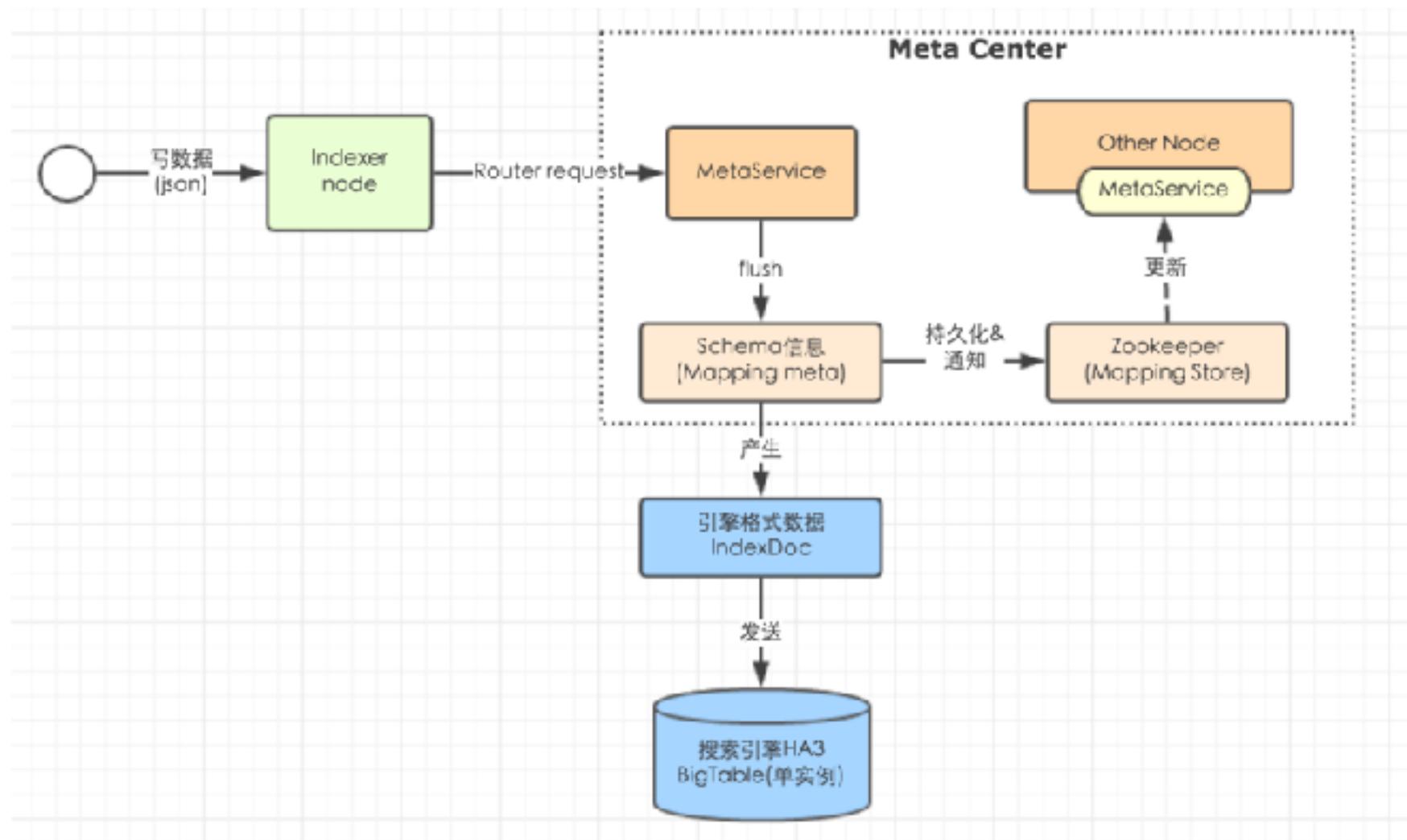
-> POST /mytable/uuid -d "{ json 数据 ... }"



Big Schema 大宽表

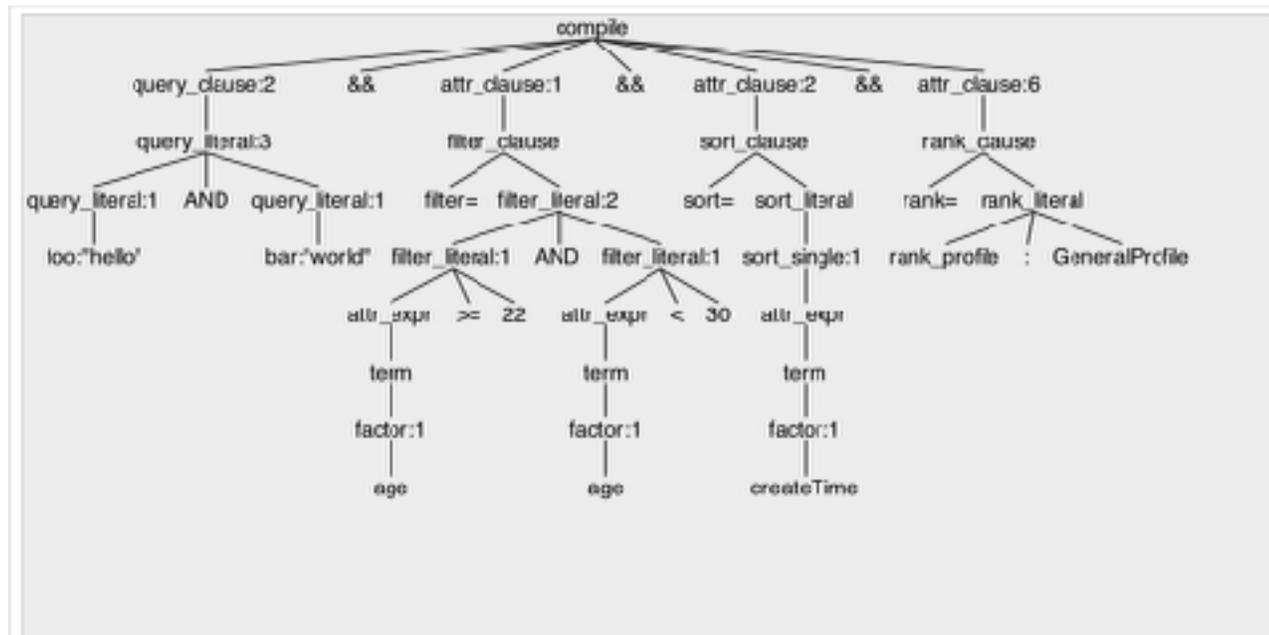
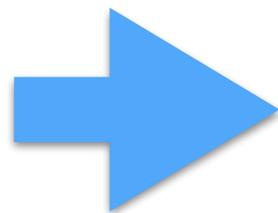


# 自动映射 - 元数据中心



# ZQL 查询语言翻译机

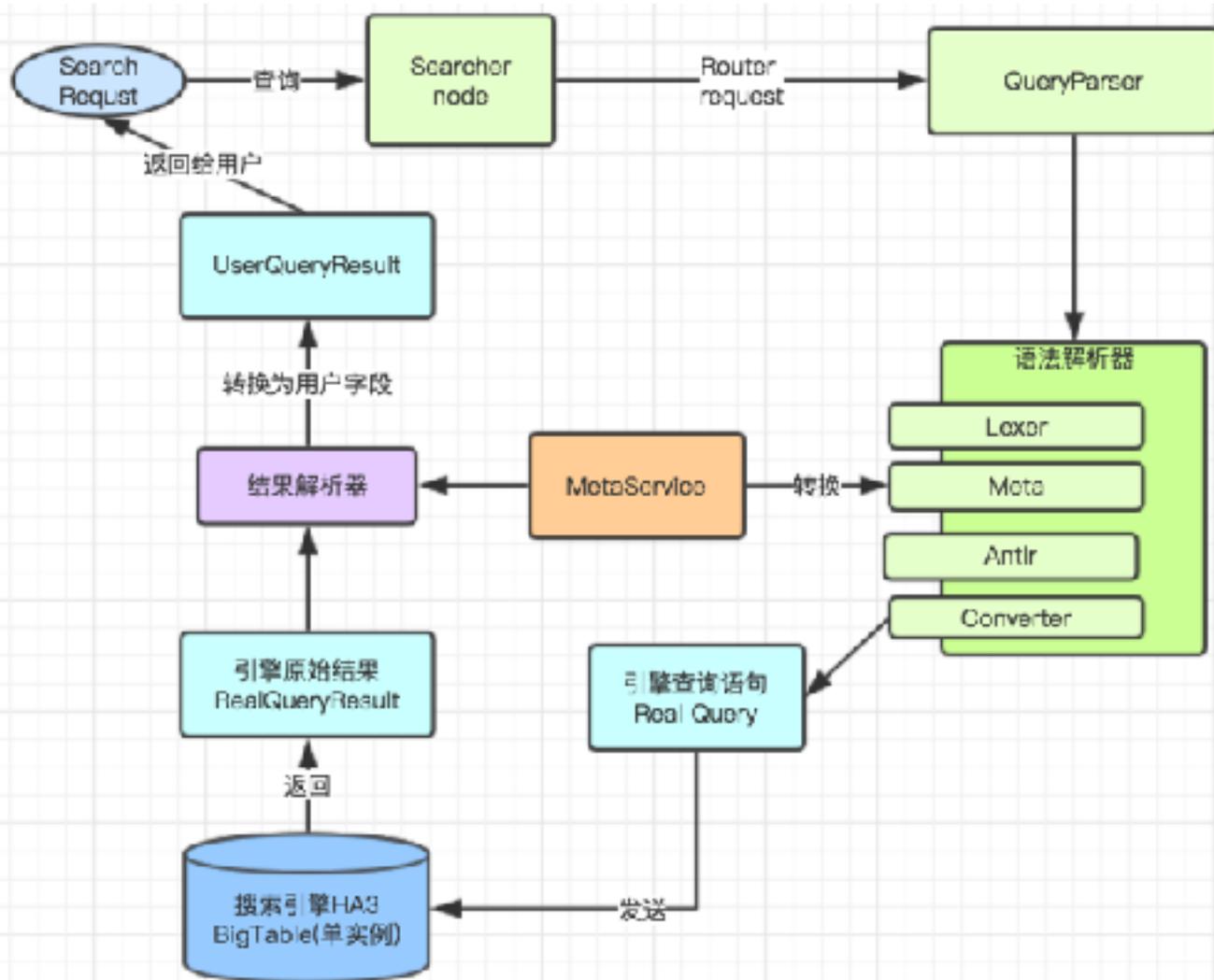
```
grammar ZQL;  
%header [...] ;  
%compile  
: qquery_clause ('&&' attr_clause) *  
; ;  
%attr_clause  
: filter_clause | sort_clause | agg_clause | distinct_clause | rank_sort_clause | rank_clause  
; ;  
%query_clause  
: 'query=' + query_literal  
| query_literal  
; ;  
%filter_clause  
: 'filter=' + filter_literal  
; ;  
%sort_clause  
: 'sort=' + sort_literal  
; ;
```



AST + Metadata



# 查询翻译过程

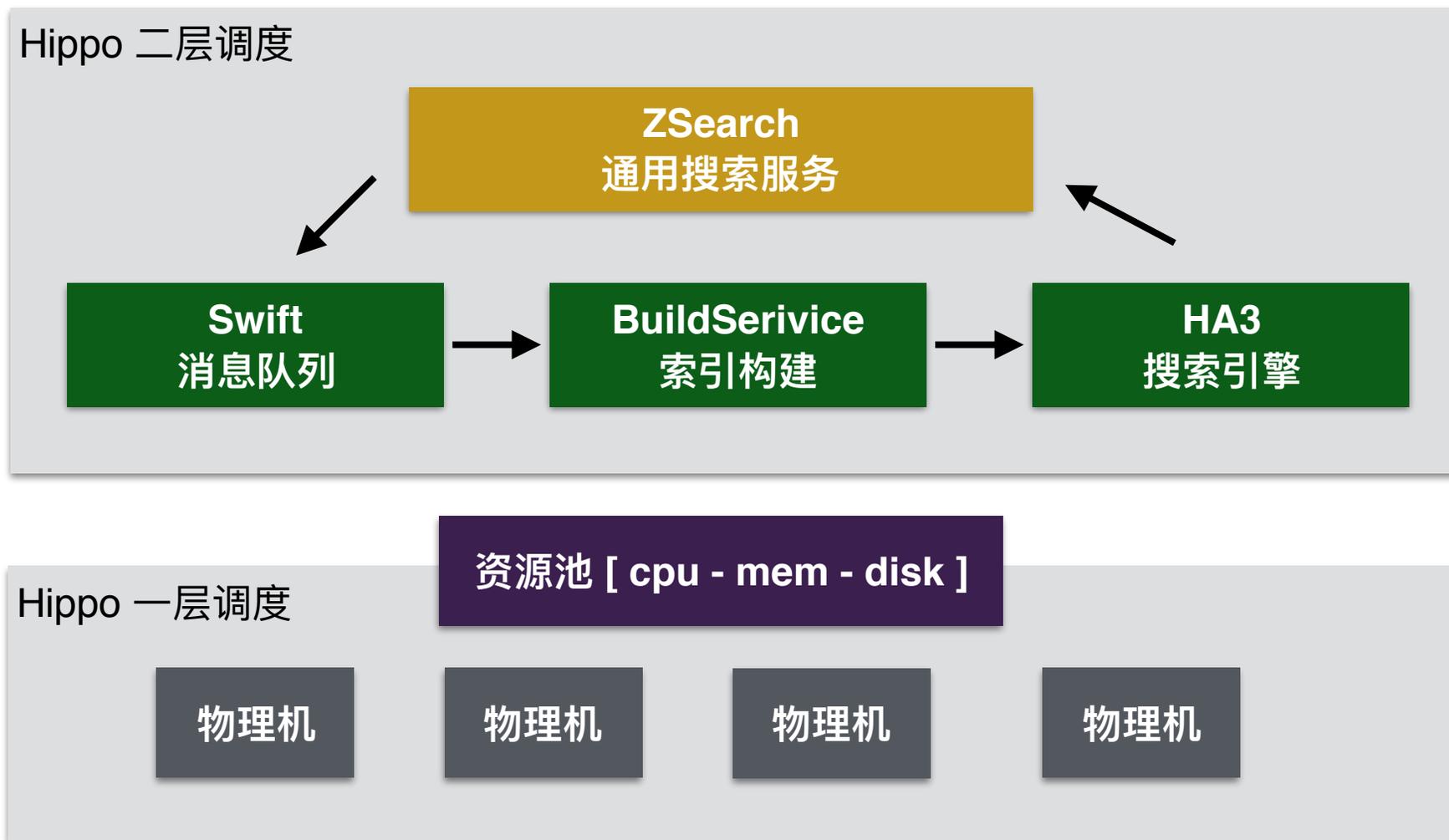


# ZSearch

- 门槛低 - 不需要任何搜索背景知识
- 易上手 - 任意格式文档、零配置、3分钟内可上手
- 性能强 - HA3引擎
- 免运维 - 云服务模式



# ZSearch 线上架构



# ZSearch 碰到的问题

- 机器成本高，不适合日志型海量数据。
- 依赖基础设施，不利于私有域的业务开展。
- 技术复杂，架构演进依赖多个团队。



# ZSearch 2.0

- 今年4月份开始引入Elasticsearch
- 良好的开源社区生态
- 架构轻量 产品成熟
- 引擎性能稍差 但成本优势大
- 没有离线部分



# ZSearch 2.0 平台定位

蚂蚁金服

阿里云

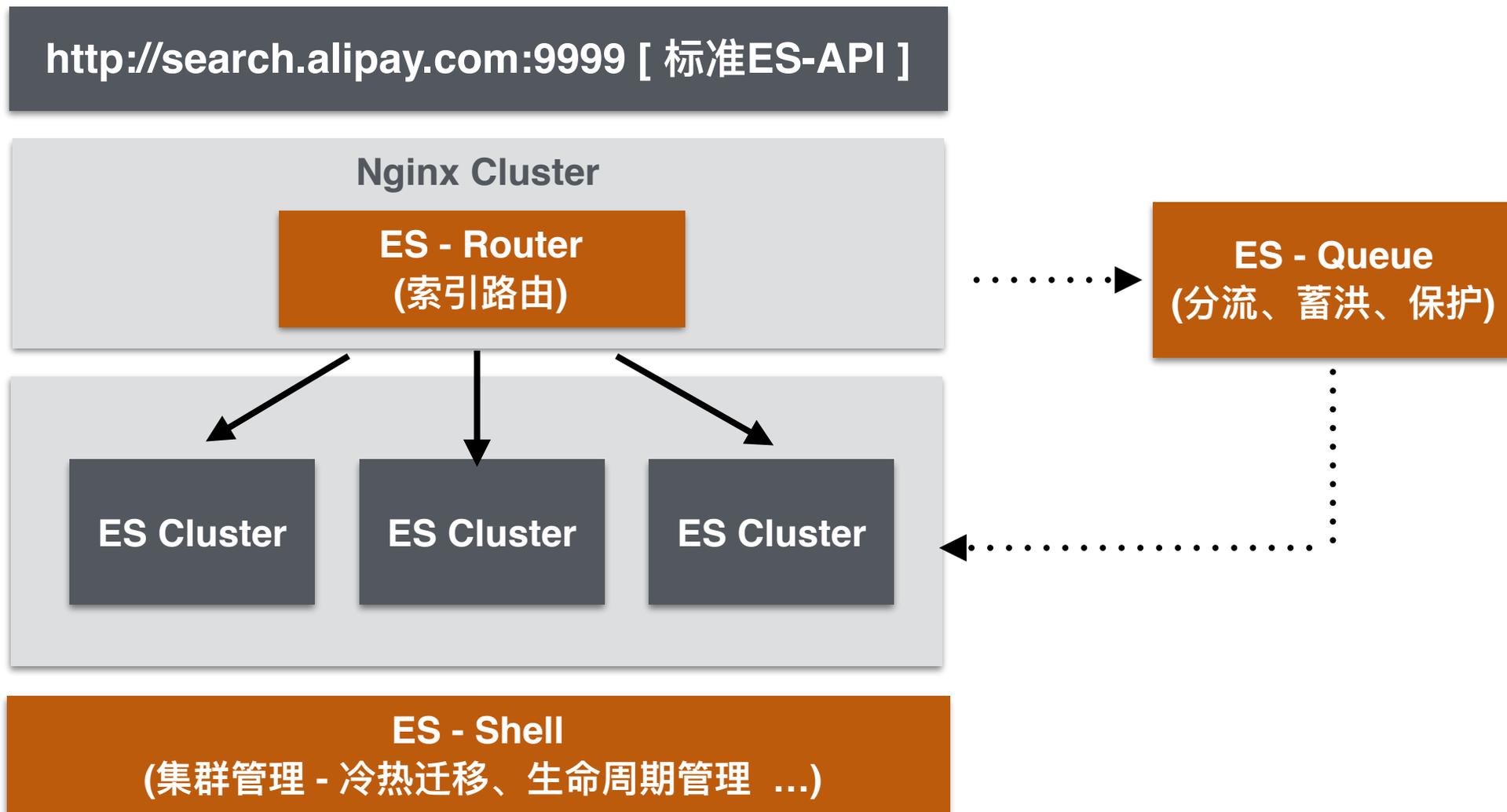
天猫

菜鸟 ..

Elasticsearch Service

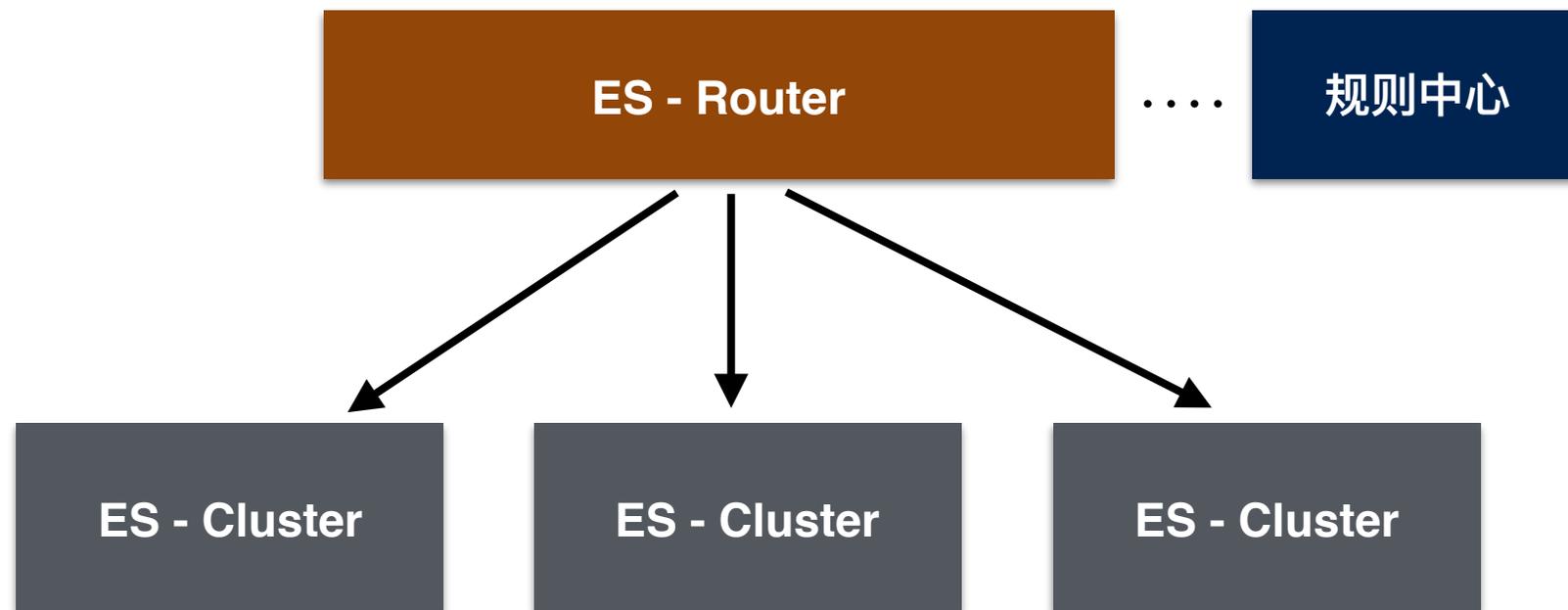


# 平台架构



# ES - Router

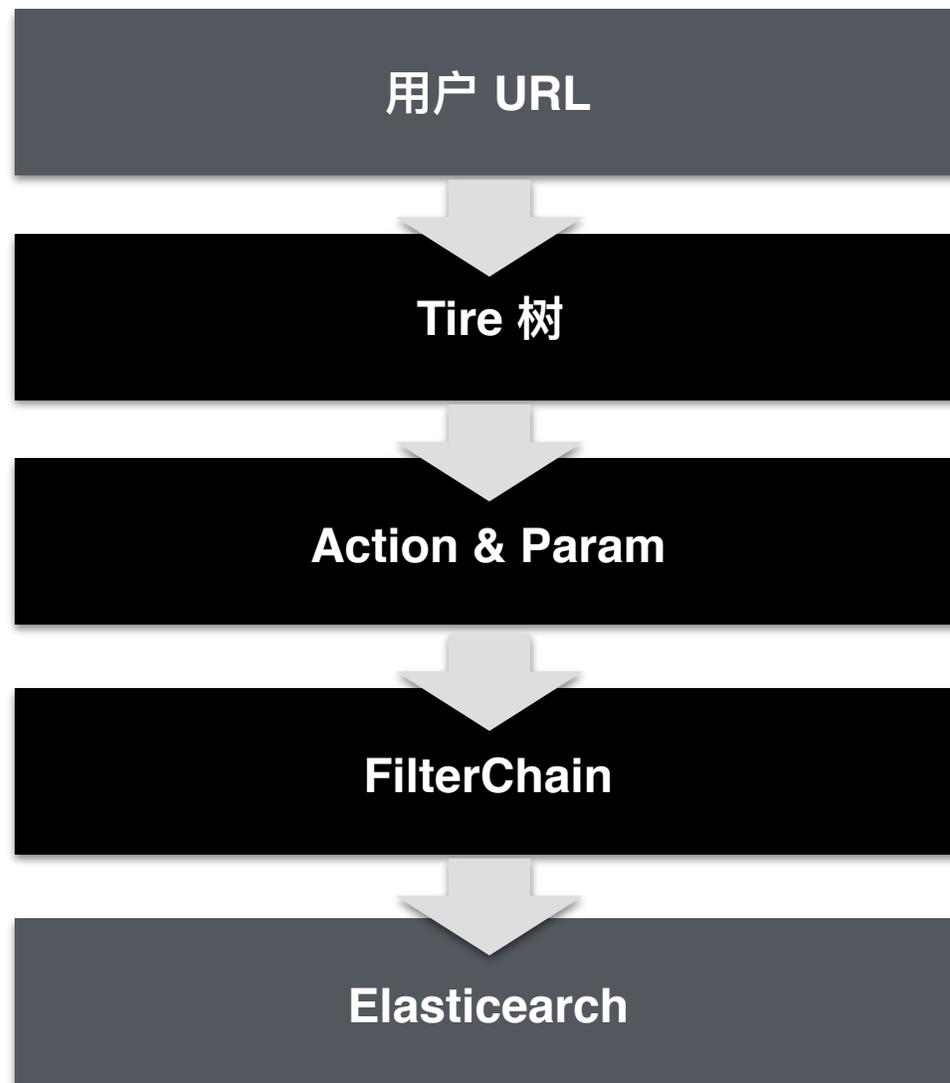
- 集群路由
- 权限隔离
- 限流保护
- 流量计费



# ES - Router 原理

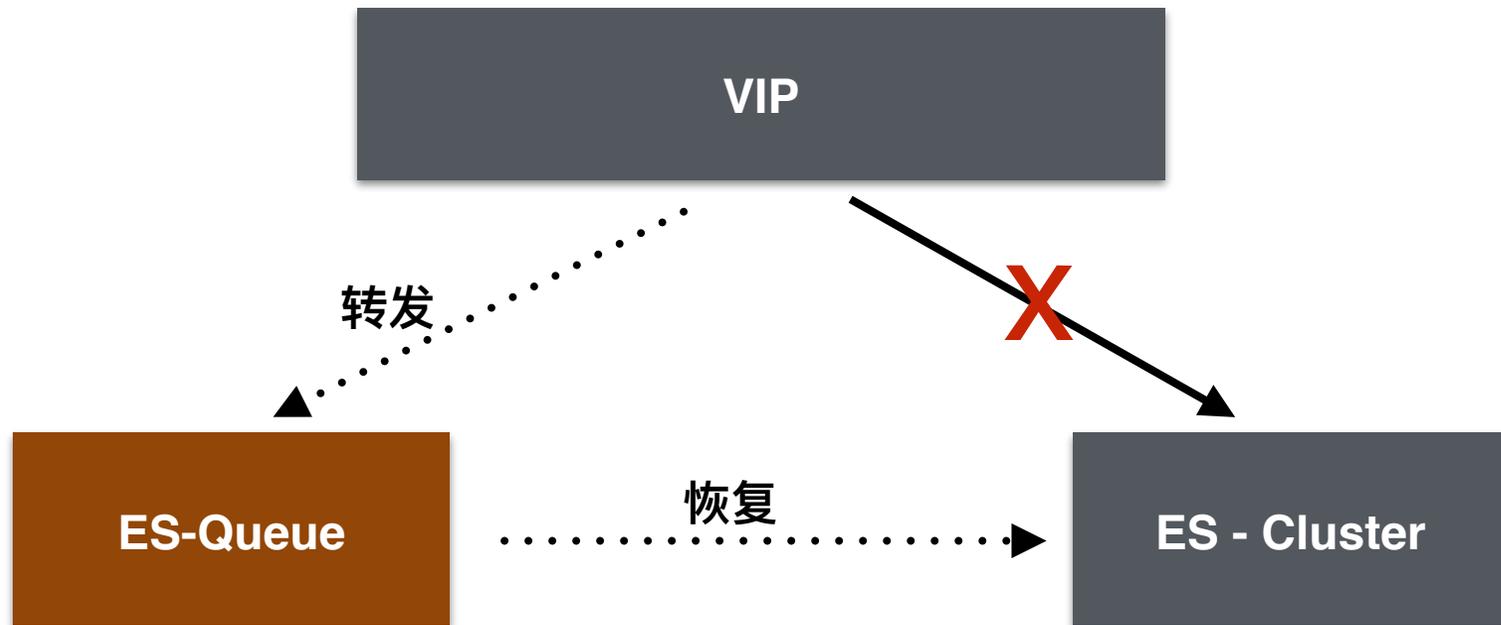
## Nginx + Lua

- 根据Tire树解析URL的ES Action
- 通过Filter添加自定义处理逻辑
- FilterChain - 路由、限流、权限 ...
- 转发到真实的ES集群



# ES - Queue

- 写入容灾
- 大促蓄洪
- 压力分流
- 在线升级



# ES - Shell

```
get_by_reg 'log.+' | filter_age_lte last_3_day |  
allocation 'warm'
```



# ES - Shell

```
POST /schedule

{
  "name": "task-foo",
  "cron": "1 * * * *",
  "description": "生命周期管理",
  "lines": [
    "get_by_reg 'zsearch.+' | filter_age_gte last_3_day | delete",
    "get_by_reg 'zsearch.+' | filter_age_gte last_2_day |
allocation 'warm' "
  ]
}
```



# Elasticsearch 三件套

- ES-Router 支持集群横向扩展，扩展平台自身能力。
- ES-Queue 实体ES集群的“充电宝”“缓冲池”。
- ES-Shell 整体索引大管家，生命周期、冷热迁移 ...

所有人看到的是一个<sup>↑</sup>原生的ES服务，保留用户习惯。



# ES 经验杂谈(1) - Http / Data / Master 隔离

网络层

Http-Node

Http-Node

Http-Node

Master-Node

数据层

Data-Node

Data-Node

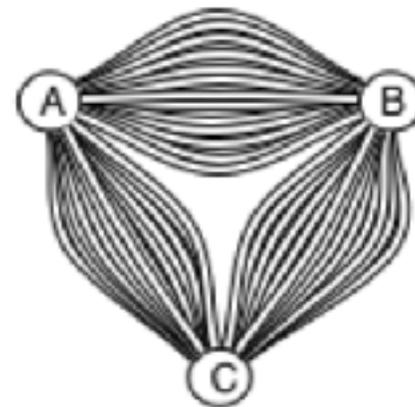
Data-Node

Data-Node



## ES 经验杂谈(2) - 集群

- ES的网络模型
- DiscoveryNode ( Local 和Remote)
- 越少的Node 性能越高
- 超过60台物理机新建集群, 用路由来横向扩展。



# ES 经验杂谈(3) - Raid

## RAID0 - 挂11块SSD

```
/dev/md0p1:
```

```
Timing cached reads: 18512 MB in 2.00 seconds = 9255.72 MB/sec
```

```
Timing buffered disk reads: 4548 MB in 3.00 seconds = 1516.00 MB/sec
```

## 单块SSD

```
/dev/sdc1:
```

```
Timing cached reads: 18486 MB in 2.00 seconds = 9252.94 MB/sec
```

```
Timing buffered disk reads: 1370 MB in 3.00 seconds = 456.50 MB/sec
```

- SSD + RAID0 性能的极致
- 副本机制保障完整性（配置物理隔离分配策略）
- RAID0 的瓶颈，受限于 memory and the card (OS, PCI)等因素。



# Q & A



# 谢谢

## THANK YOU

---

城破

