

基于ES的时序数据库服务

johngqiang (姜国强)

目录



elastic
中文社区

IT大咖说
知识共享平台

CFD基础架构部

- 背景
- 功能特性
- 竞品对比
- 业务现状
- 遇到的问题

一、背景

什么是时间序列数据？

- 按时间顺序记录系统、设备状态变化的数据
- 典型场景：
 - DevOps监控
 - 应用程序指标
 - IoT传感器



一、背景



elastic
中文社区

IT大咖说
知识共享平台

CFD基础架构部

什么是时序数据库（TSDB）？

- 针对时序数据的特点对写入、存储、查询进行优化的专业数据库

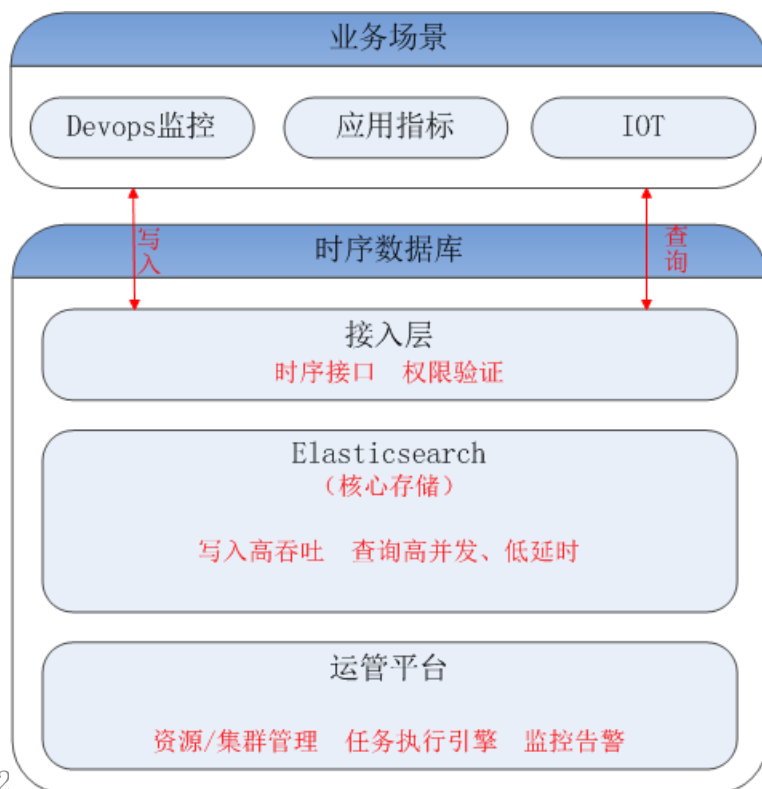
常见特性

- 高并发写入能力
- 针对维度进行索引，优化查询
- 高效压缩，降低存储成本
- 数据生命周期管理
- 优化长期存储，降低成本【高级】
- 支持数据复杂的聚合、降精度等【高级】

二、功能特性

基于Elasticsearch构建

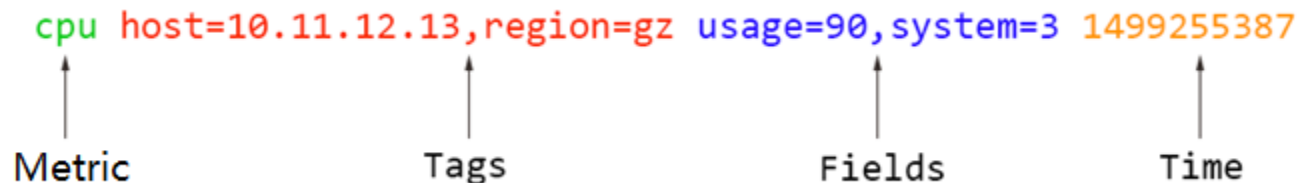
- 高写入性能、多维分析能力
- 集群化，易使用，维护成本低



二、功能特性

时序模型

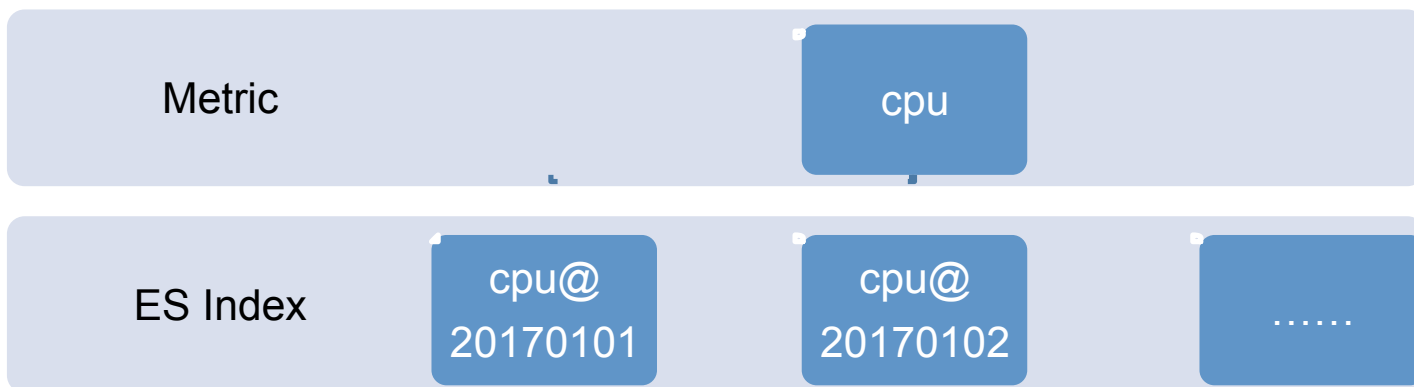
- 与常用时序模型保持一致
- 多指标模型



术语	说明	关系型数据库对照
Metric	业务关注的一份数据集合	Table
Tags	维度列，可指定条件搜索	具有索引的Column
Time	时间列，一个特殊的维度	具有索引的Column
Fields	指标列，数值随时间变化	不具有索引的Column

二、功能特性

时序模型

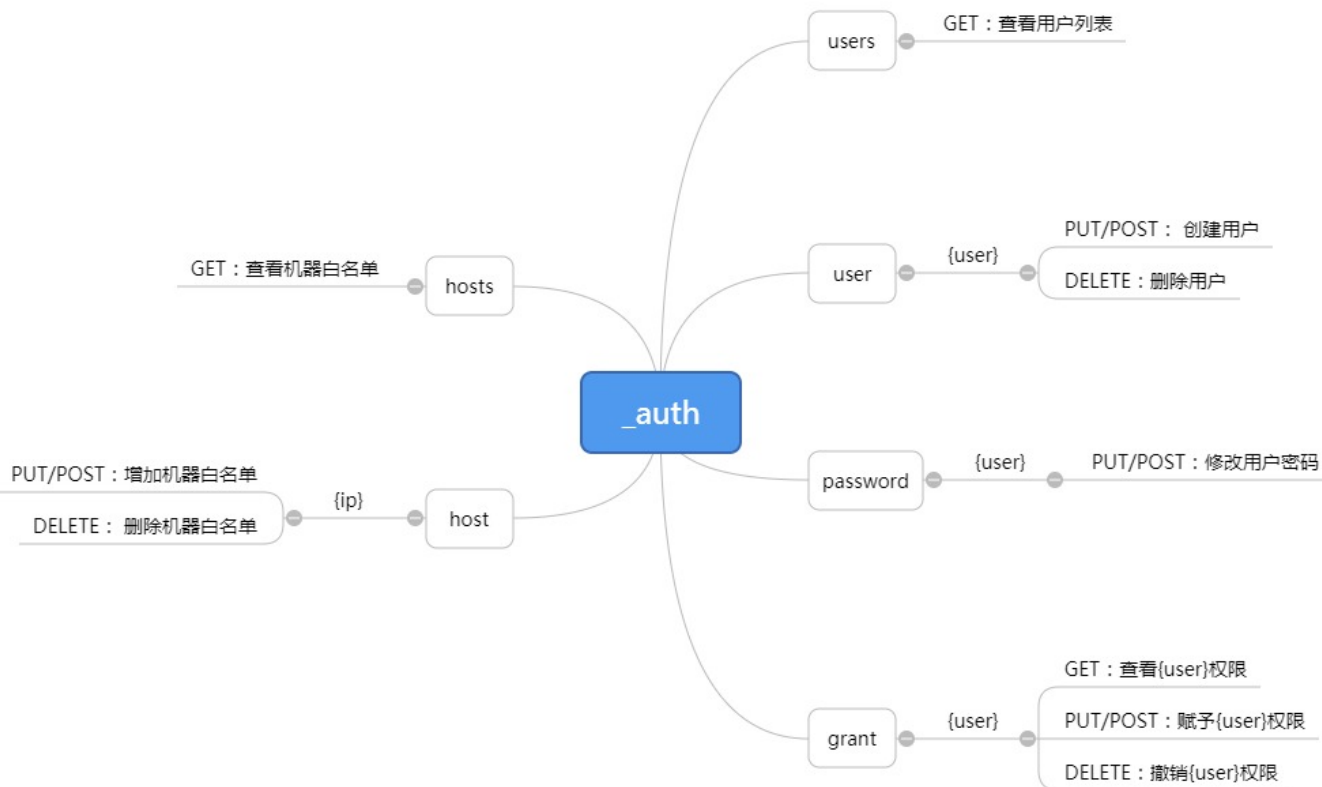


```
1 POST /cpu/_bulk
2 {"index":{}}
3 {"region":"gz","host":"10.12.13.14","time":1499255387,"usage":90,"system":3}
4 {"index":{}}
5 {"region":"sh","host":"14.13.12.10","time":1499255387,"usage":50,"system":1}
6
7 POST /cpu/_search
```


二、功能特性

权限系统

- 业务对数据访问有安全性需求，要求读写分离
- 支持Http Base Auth认证、机器白名单



二、功能特性

权限系统

- Elasticsearch REST接口风格
- 性能影响非常低：1%~

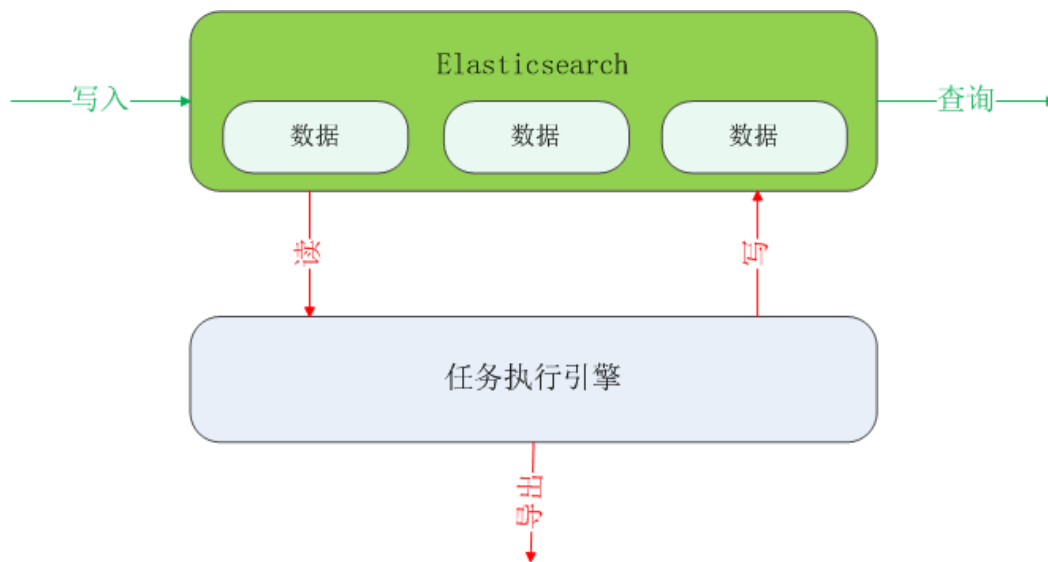
```
1 POST _auth/user/test_user
2 {
3   "password": "test_password"
4 }
5
6 POST _auth/grant/test_user
7 {
8   "index": "test_index_*",
9   "privileges": "read,write"
10 }
11
12 PUT _auth/password/test_user
13 {
14   "password": "test_password_new"
15 }
16
17 POST _auth/host/192.168.0.1
18
```

	平均写入速度 (条/秒)	平均查询并发数 (次/秒)
未安装权限系统	224636	28100
安装权限系统	220759	27950

二、功能特性

任务执行引擎

- 集群内部复杂的、离线的任务调度及执行
 - 数据导出：过滤结果、样例数据
 - 数据降精度处理
 - 监控数据采集

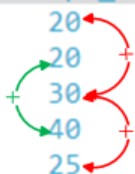


二、功能特性

降精度（预聚合）

- 降低时间维度精度，保留粗粒度历史数据
 - 进一步加快长时间跨度查询的性能
 - 降低历史数据存储成本

1	region	host	time	cpu_usage
2	gz	10.11.12.13	2017-07-01 10:00:00	20
3	bj	14.15.16.17	2017-07-01 10:00:00	20
4	gz	10.11.12.13	2017-07-01 10:00:10	30
5	bj	14.15.16.17	2017-07-01 10:00:10	40
6	gz	10.11.12.13	2017-07-01 10:00:20	25



1	region	host	time	cpu_usage
2	gz	10.11.12.13	2017-07-01 10:00:00	25
3	bj	14.15.16.17	2017-07-01 10:00:00	30

二、功能特性

数据生命周期管理

- Index按保留时间、写入速度自动滚动
- 数据超出过期时间后自动清理

保存表 ×

表名

简介

过期时间 天 ▼

二、时序数据库

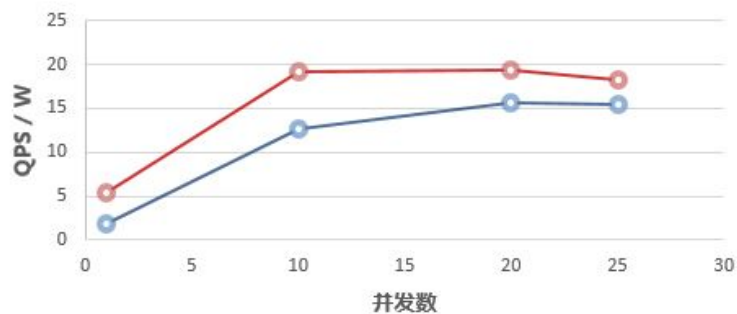
测试工具

- <https://github.com/influxdata/influxdb-comparisons>
- 说明：ES测试部分修复部分Bug及调优

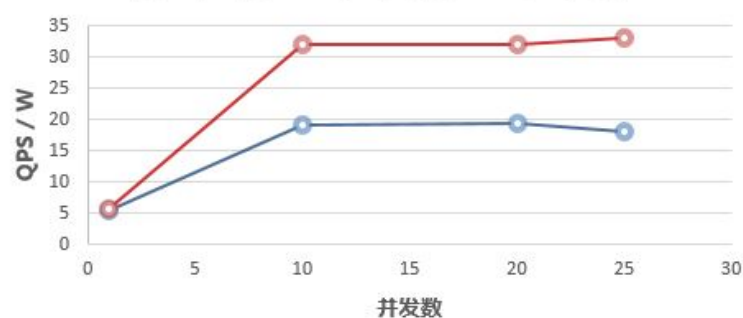
写入测试

- ES单机写入能力19w/s，略优于InfluxDB
- ES具有近似线性扩展的分布式方案

写入性能 InfluxDB-ES单节点



写入性能 ES单节点-ES双节点



—●— InfluxDB吞吐/w —●— es单节点吞吐/w

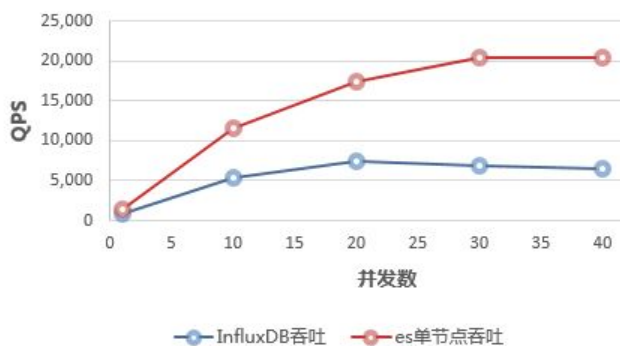
—●— es单节点吞吐/w —●— es双节点吞吐/w

二、时序数据库

查询测试

- 开启routing时，单机ES查询性能为2w，接近InfluxDB的4倍
- 无routing时二者性能接近
- ES查询具有线性扩展能力，没有时间线、单维度唯一值上限限制

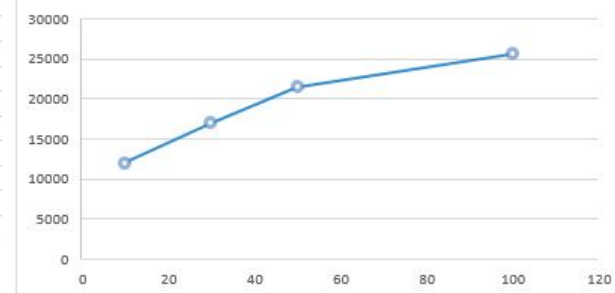
1host读性能 InfluxDB-ES单节点



1host读性能 ES单节点-ES双节点



50w 实例数查询性能测试



三、竞品对比

	ES	InfluxData	Prometheus	Graphite	OpenTSDB
数据模型	labels	labels	labels	dot -separated	labels
写入性能	*****	*****			***
压缩编码	****	*****			*****
读取性能	*****	****			****
数据生命周期管理	✓	✓	✓	✓	手动
集群化支持	✓	√商业版	单机	单机	✓
降精度（预聚合）	✓	✓	×	✓	×
权限管理	✓	√商业版	×	×	×

三、竞品对比



elastic
中文社区

IT大咖说
知识共享平台

CFD基础架构部

	ES	InfluxData	Prometheus	Graphite	OpenTSDB
外部依赖	无	无	无	采集、内部任务依赖	Hadoop & HBase
接口	REST	类SQL	REST	REST	REST
社区生态	+++++	+++	++	++	++
聚合分析	强	弱	弱	弱	弱
延伸应用	日志、全文检索等场景	×	×	×	×

四、业务现状



elastic
中文社区

IT大咖说
知识共享平台

CFD基础架构部

总体部署

- 共部署**400+**台机器，600+个ES节点
- 部署超过10+地域
- 支持腾讯内部20+业务
 - 云监控、云数据库、云负载、财付通、彩票等

最大单集群

- 50台机器，启动150个ES节点
- 写入流量：峰值QPS在**300w/s**，每天20TB+

五、遇到的问题



elastic
中文社区

IT大咖说
知识共享平台

CFD基础架构部

1. 大集群请求Hang

- 现象：(150 Nodes)
 - 有一定比例的访问请求Hang不返回
 - 多节点内存逐渐升高后OOM，集群崩溃
- 复现方式：
 - 在3台物理机上搭建150 Nodes的集群
 - 重启其中一个Node
 - 给所有Node发送请求，部分节点可能Hang
- 问题原因：大集群在进行节点间通信时，容易导致tcp backlog queue打满，而5.6.3之前版本会复用有问题的连接
- ISSUE：<https://github.com/elastic/elasticsearch/issues/25863>

五、遇到的问题

2. 大量Shard.....

- 背景：
 - 腾讯云监控是统一监控平台，支持众多不同类型的监控需求
 - Shard数量暴涨：1000 Index * 31天 * 5 Shard * 2副本
- 问题：
 - 分片过多后，建Index非常慢（分钟级）
 - 凌晨集中建表，建表速度慢，写入速度快，拖垮集群
- 解决：
 - 依据数据保留时长滚动Index
 - 分散、提前创建Index
 - 结合分片大小、写入速度分配Shard数量

时序数据库即将开启

极高性能写入、查询、聚合，天生适用IoT、日志存储等场景

立即预约

QA

欢迎加入我们