

The background is a solid blue color with decorative network-like graphics in the corners, consisting of interconnected nodes and lines. The Gdevops logo is centered at the top, with a white 'G' and 'devops' in white text.

Gdevops

全球敏捷运维峰会

携程机票Elasticsearch集群驯服记

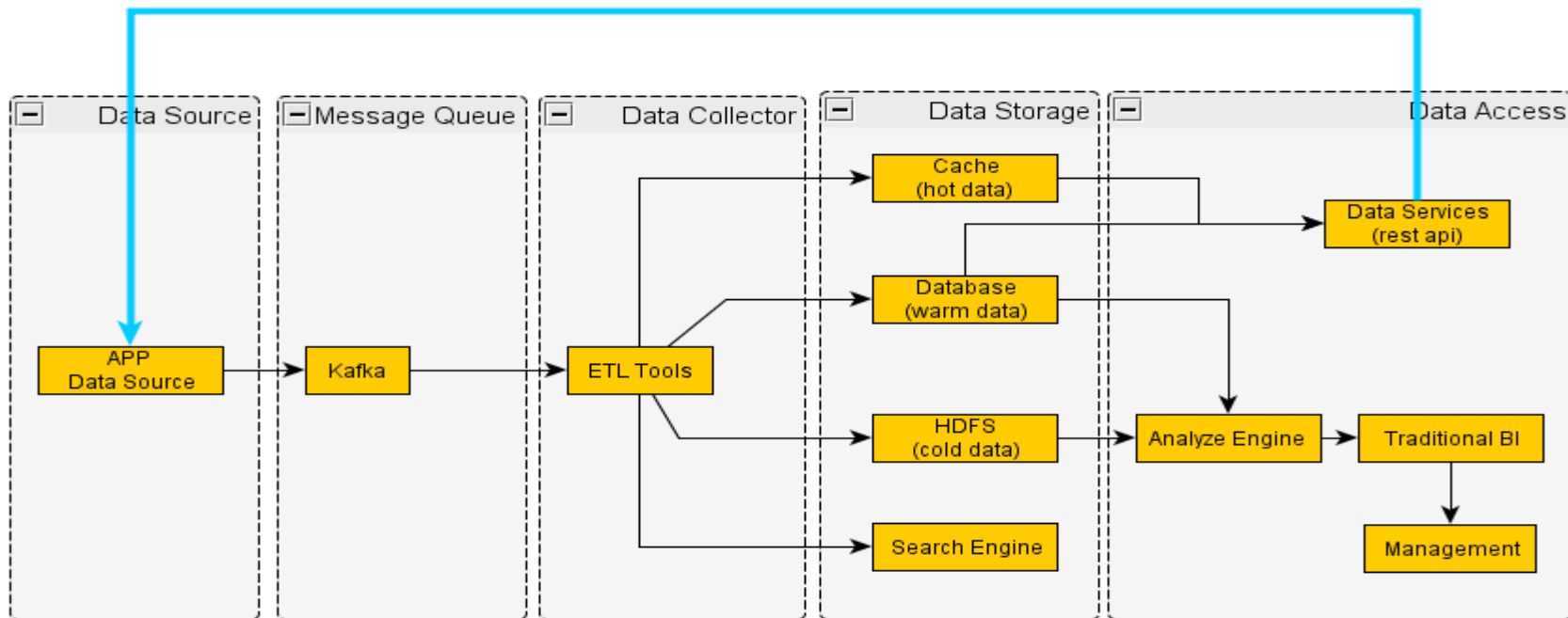
演讲人：许鹏



目录

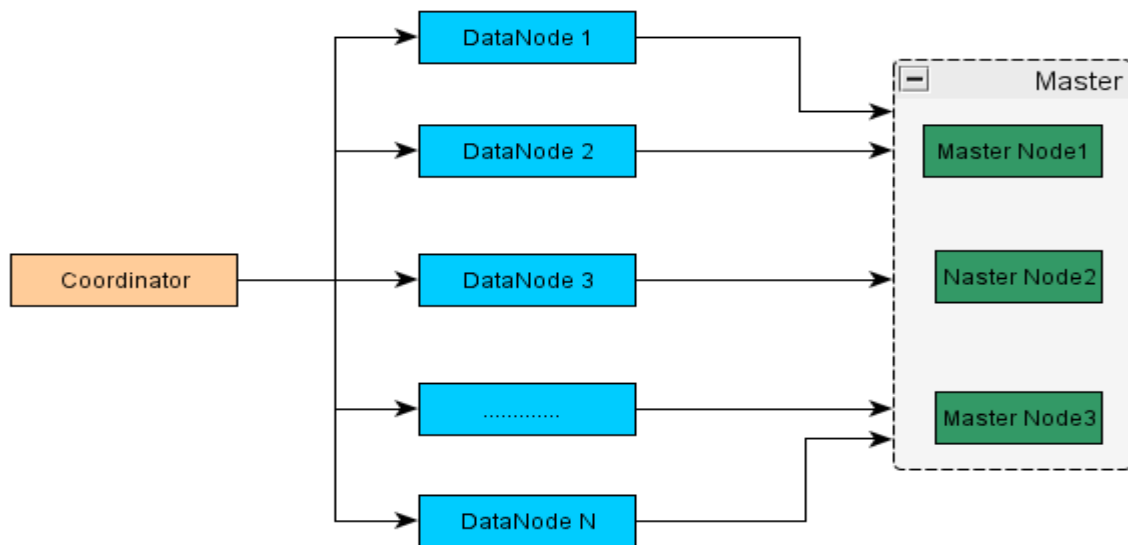
- 集群规划
- 集群设置
- 集群监控

整体架构



集群规划

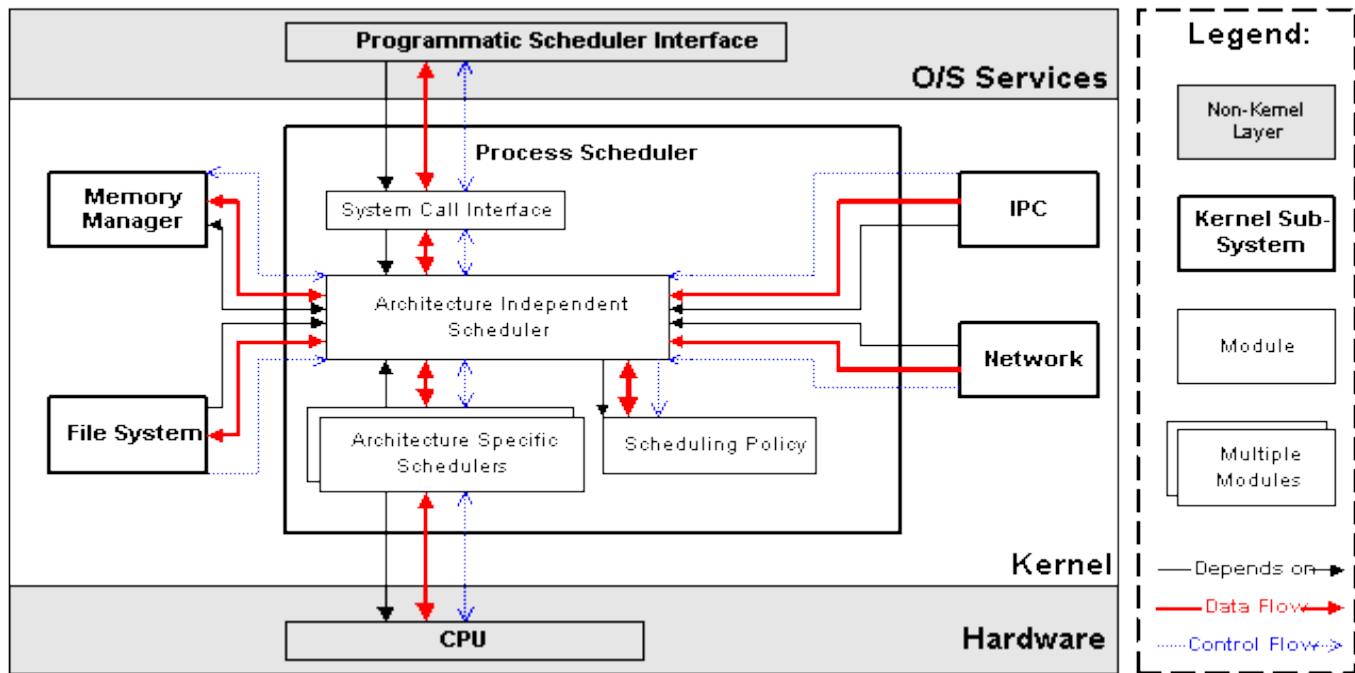
- Coordinator Node
 - 协调节点
 - 不进行数据存储
- Master Node
 - 集群节点信息
 - 索引元信息
- Data Node
 - 数据存储节点





PART II 集群配置

底层操作系统



OS参数设置 - 内存相关

- */etc/security/limits.conf*

- * hard memlock unlimited
- * soft memlock unlimited
- * - nofile 65535

- 确保 **/etc/pam.d/login** 文件中有如下内容

```
session required /lib/security/pam_limits.so
```

- `max_map_count` 定义了进程能拥有的最多内存区域

```
sysctl -w vm.max_map_count=262144
```

OS参数设置 - IO相关

1. /etc/fstab

```
/dev/sda1          /opt/data/1      ext4 defaults,noatime,nodiratime 0 0  
/dev/sdb1          /opt/data/2      ext4 defaults,noatime,nodiratime 0 0
```

noatime, nodiratime避免每次数据访问的时候都更新access time信息

2. 设置 vm.dirty_ratio和vm.dirty_background_ratio

```
sysctl -w vm.dirty_ratio=10
```

```
sysctl -w vm.dirty_background_ratio=5
```

3. 设置swap

```
sudo sh -c 'echo "1">/proc/sys/vm/swappiness'
```

4. ioscheduler 如果是ssd硬盘，建议使用deadline

```
sudo sh -c 'echo "cfq">/sys/block/sdc/queue/scheduler'
```


Elasticsearch参数设置 – JVM设置

1. /etc/elasticsearch/jvm.options

-Xms32g

-Xmx32g

-XX:+ExitOnOutOfMemoryError

注: *ExitOnOutOfMemoryError*从**jdk 1.8.0_92**开始支持, 如果是低版本可以使用-
XX:OnOutOfMemoryError="kill -9 %p"

2. /etc/elasticsearch/elasticsearch.yml

bootstrap.memory_lock: true

bootstrap.system_call_filter: false

Elasticsearch 参数设置

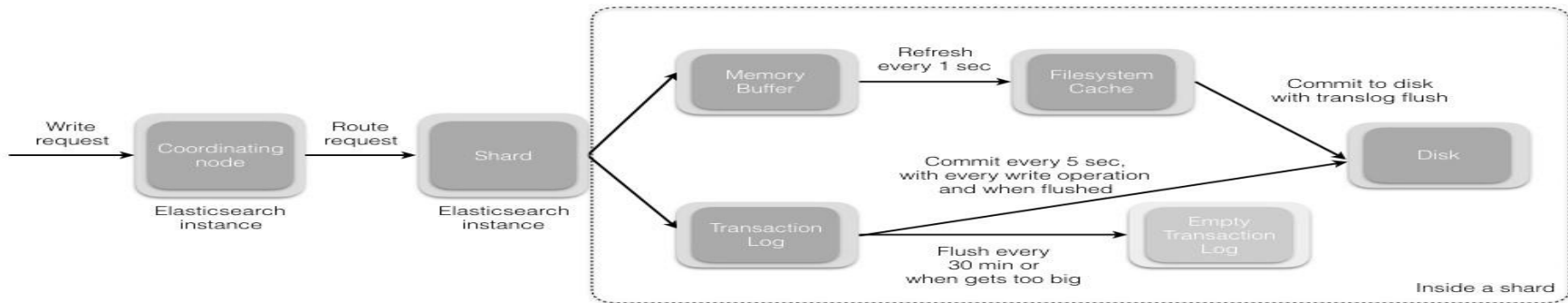
- Shard 均匀分布
- Recovery

	shop_de_v1 shards: 5 * 3 docs: 169119 size: 54.69MB	shop_nl_v1 shards: 5 * 3 docs: 131858 size: 33.57MB	shop_pl_v1 shards: 5 * 3 docs: 130089 size: 37.16MB
★ Spielwiese inet: [redacted] Xz1v1vexStqP9uISuCu1nQ load: 0.00 heap: 222.97MB/332.25MB	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
☆ nl-1 inet: [redacted] x099CnMSBCWmXL8Y12uBA load: 0.00 heap: 1.52GB/7.97GB	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
☆ PL-1 inet: [redacted] mEso0CnRvZhtD50wVudnQ load: 0.00 heap: 95.27MB/3.98GB	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
🔊 unassigned shards			

PUT _cluster/settings

```
{
  "transient": {
    "cluster": {
      "routing": {
        "allocation": {
          "cluster_concurrent_rebalance": "4",
          "node_concurrent_recoveries": "2",
          "exclude": {
            "_name": "",
            "_ip": "192.168.0.111"
          }
        },
        "balance": {
          "index": "2.0f",
          "shard": "0.2f"
        }
      },
      "enable": "all"
    }
  }
}
```

索引参数设置



名称	推荐值	
index.routing.allocation.total_shards_per_node	3	
index.refresh_interval	15s	
index.number_of_shards	12	
index.number_of_replicas	1	
index.store.type	mmapfs	
index.translog.flush_threshold_size	2g	
index.merge.scheduler.max_thread_count	1	

索引 Mapping 设置

Dynamic Mapping	<pre>"dynamic_templates": [{ "string_template": { "match_mapping_type": "string", "mapping": { "ignore_above": "10915", "type": "keyword" } }]</pre>
Object	<pre>{ "field_name": { "type": "object", "enabled": false } }</pre>

统一调用接口

- 不直接开放Elasticsearch的查询，而是通过开发的rest api来支持
 - 可以对所有的查询进行监控
 - 降低学习曲线
 - Elasticsearch-SQL
- 不足
 - Elasticsearch-SQL 能够处理的分析函数有限



Part III 集群监控

监控内容

- OS层面

- 内存
- CPU

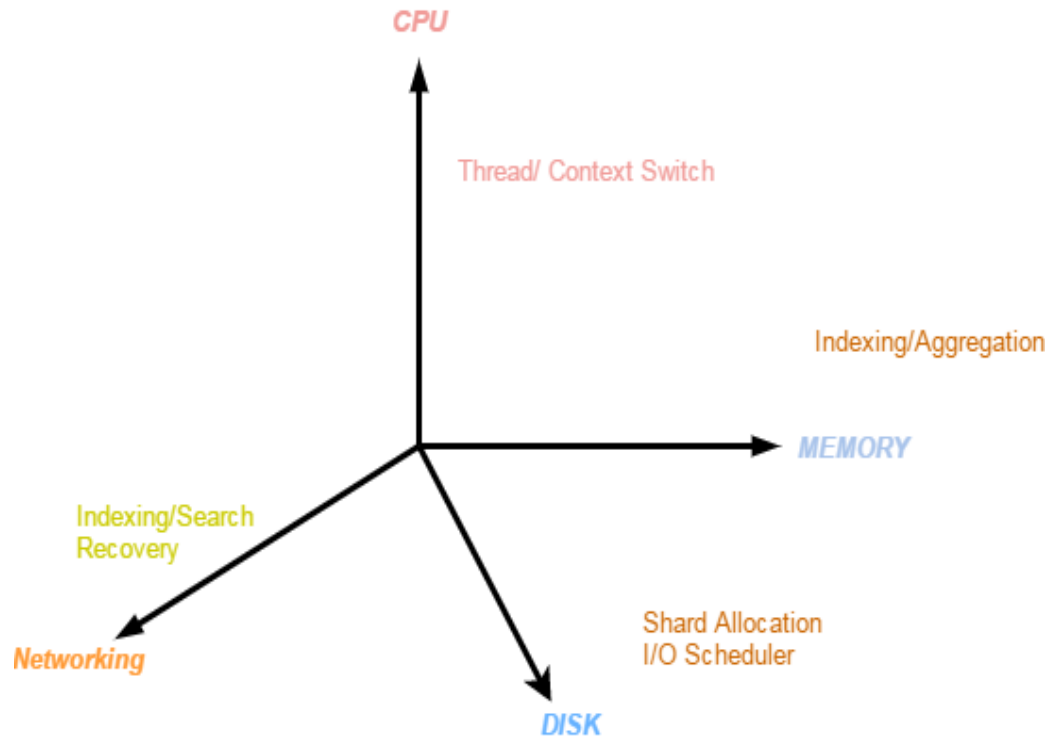
- 索引层面

- Shard
- Field Data
- 占用内存

	监控	
索引	Shard数	
	Shard分布均匀	
	是否存在大量查询	
	避免mapping中字段过多，引起mapping的频繁更新	
	聚合操作常易引起OOM	

监控工具

- elasticsearch.log
 - 节点退出
 - 内存溢出
- /var/log/messages
 - 网络错误
 - 硬件故障
- 多用_cat api
- X-Pack
- Eyeones
 - <https://github.com/hseagle/eyeones>



CAT API

集群状态

GET _cat/health
GET _cluster/health?pretty
GET _cluster/state

索引信息

GET _cat/indices

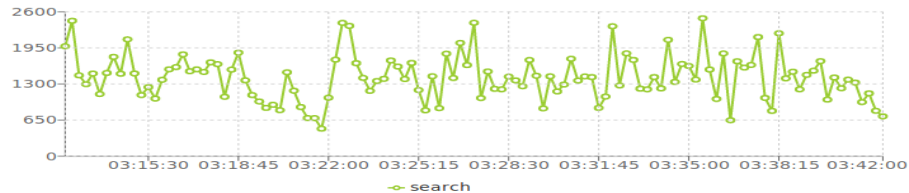
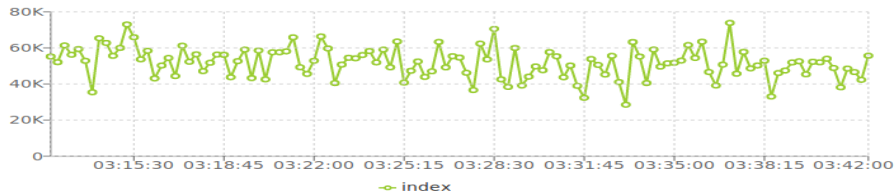
节点状态

GET _cat/nodes
GET _nodes/stats

shard信息

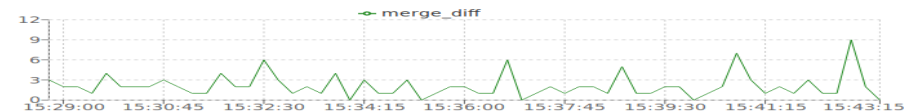
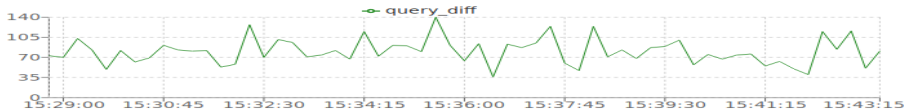
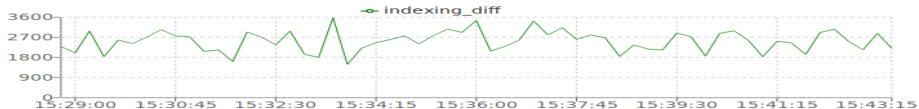
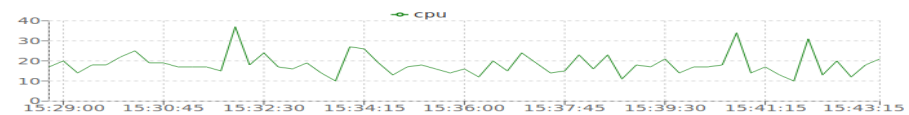
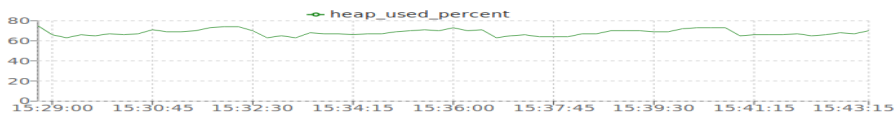
GET _cat/shards

监控内容 -- 集群和节点

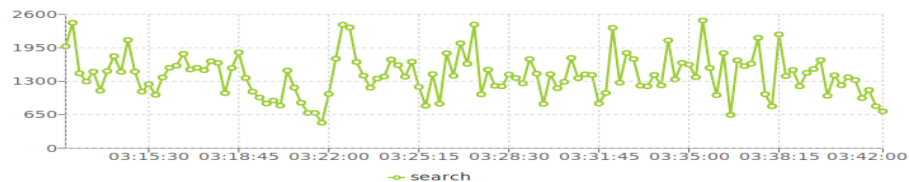
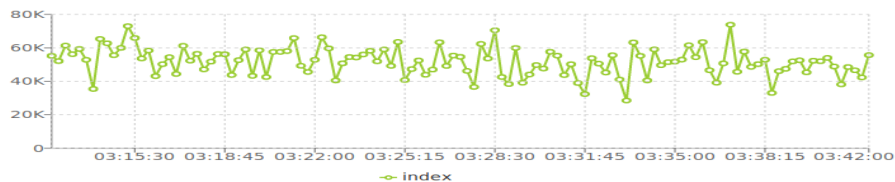


name	load_5m	heap_used_percen...	cpu	ip	index	query	shard_num...	available_d...	fielddata
[blurred]	18.4	74	14	[blurred]	4126	122	534	37TB	77MB
[blurred]	17.88	67	31	[blurred]	2415	107	535	17TB	67MB
[blurred]	16.86	72	38	[blurred]	3057	102	535	14TB	52MB
[blurred]	16.53	74	25	[blurred]	3306	107	535	19TB	69MB

Search
[blurred]
[blurred]
[blurred]
[blurred]

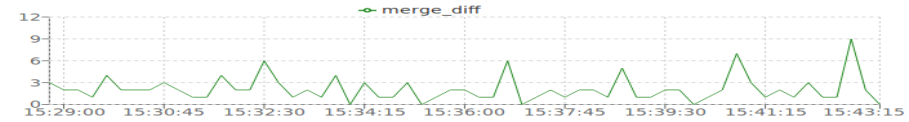
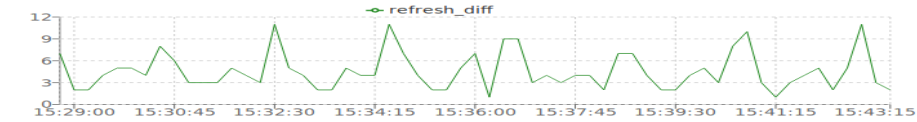
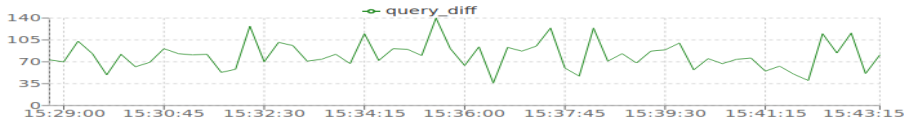
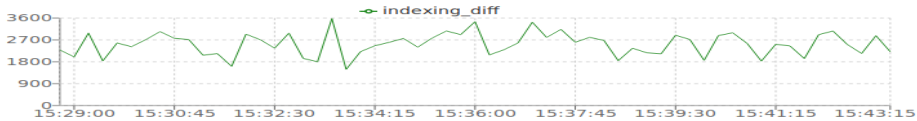
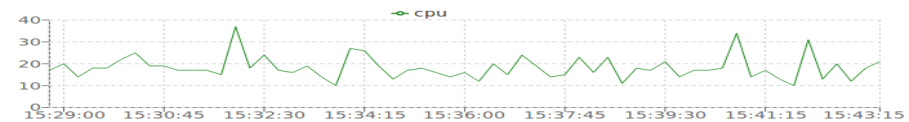
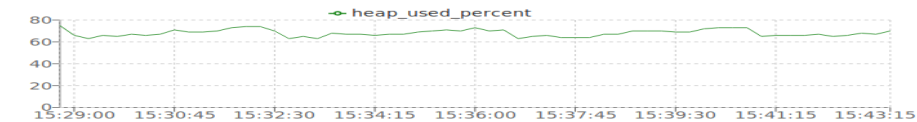


监控内容 -- 集群和节点



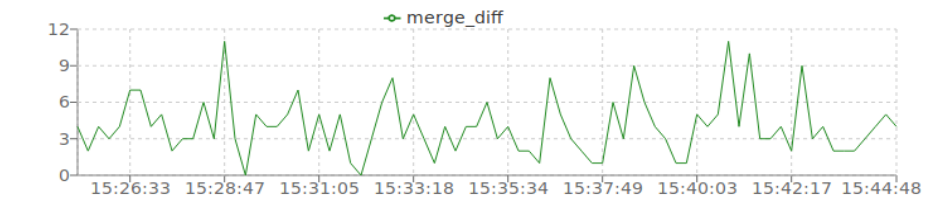
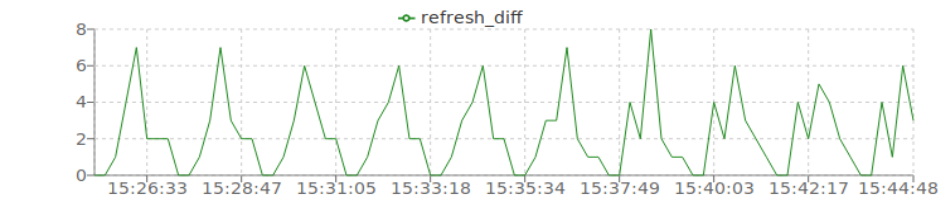
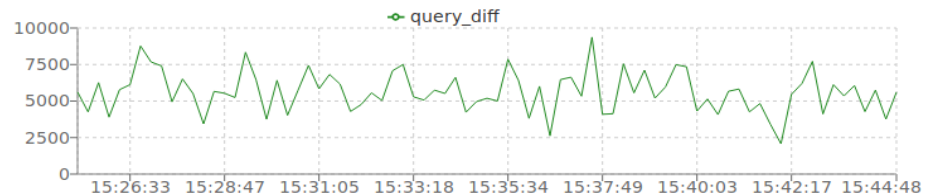
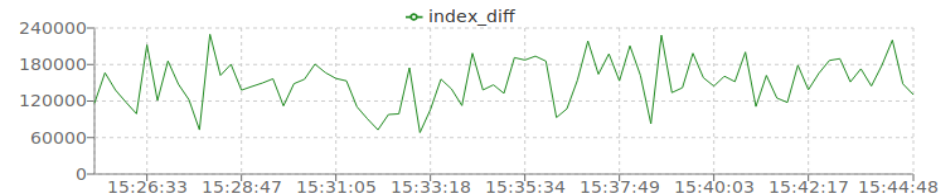
name	load_5m	heap_used_percent	cpu	ip
...	18.4	74	14	...
...	17.88	67	31	...
...	16.86	72	38	...
...	16.53	74	25	...


Search	index	query	shard_num...	available_d...	fielddata
...	4126	122	534	37TB	77MB
...	2415	107	535	17TB	67MB
...	3057	102	535	14TB	52MB
...	3306	107	535	19TB	69MB



监控内容 - 索引状态

name	index	query	docs	refresh	merge	shard_num	size
[REDACTED]	9644	404	333414432	0	2	18	456GB
[REDACTED]	7707	302	139848074	1	0	36	450GB
[REDACTED]	6149	16	148835435	5	1	18	574GB
[REDACTED]	4127	0	196980906	0	0	24	23GB
[REDACTED]	2366	0	173676105	0	3	18	250GB
[REDACTED]	1409	0	63426570	1	1	12	273GB



- 
- 深度课程: Elasticsearch从原理到实战
 - 基本概念
 - 安装与部署
 - 查询和分析
 - 性能监控与调优
 - 应用开发

The logo for Gdevops, featuring a stylized orange 'G' followed by the word 'devops' in white lowercase letters. The background is blue with decorative geometric patterns and network-like structures in the corners.

Gdevops

全球敏捷运维峰会

THANK YOU!