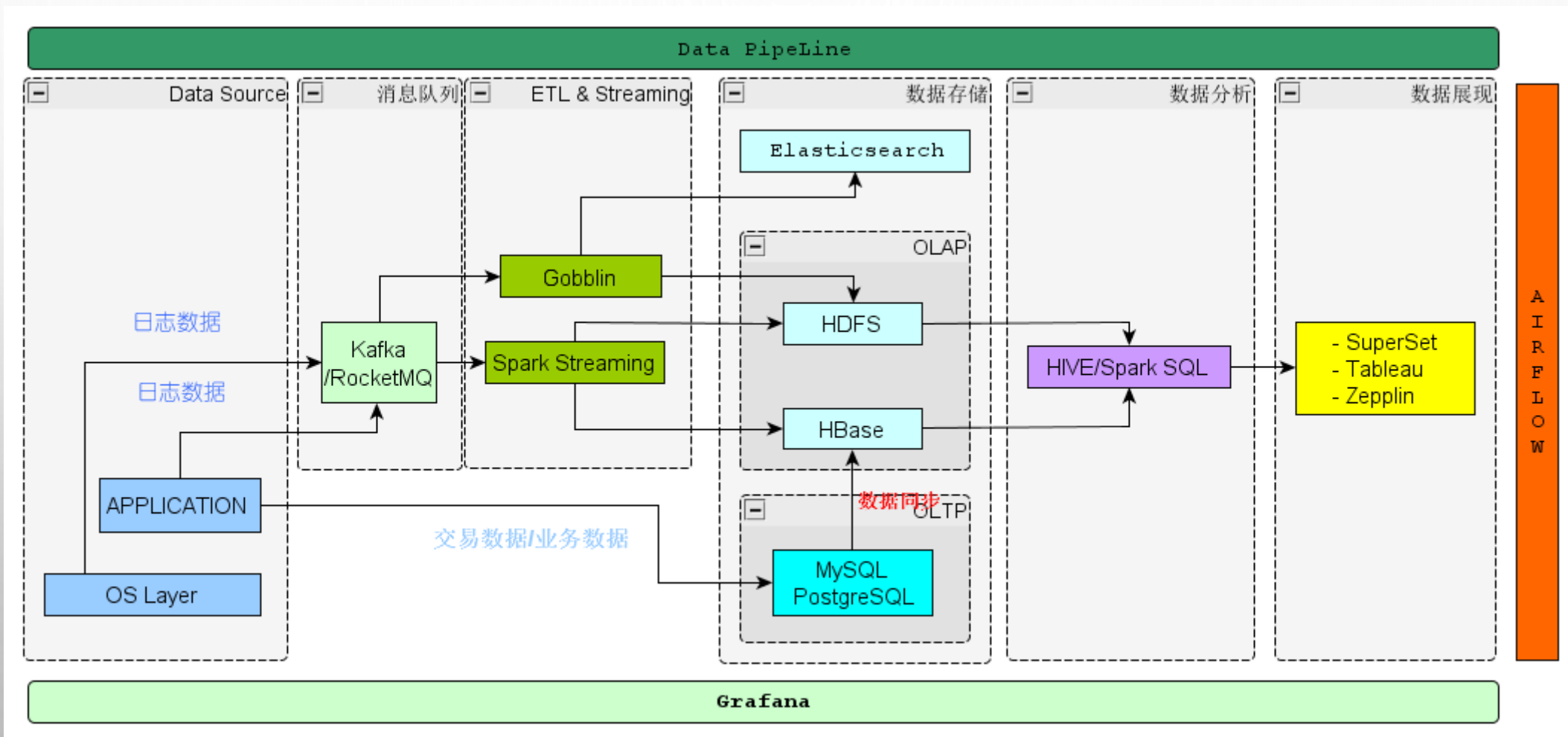


# ELASTICSEARCH集群运维实战

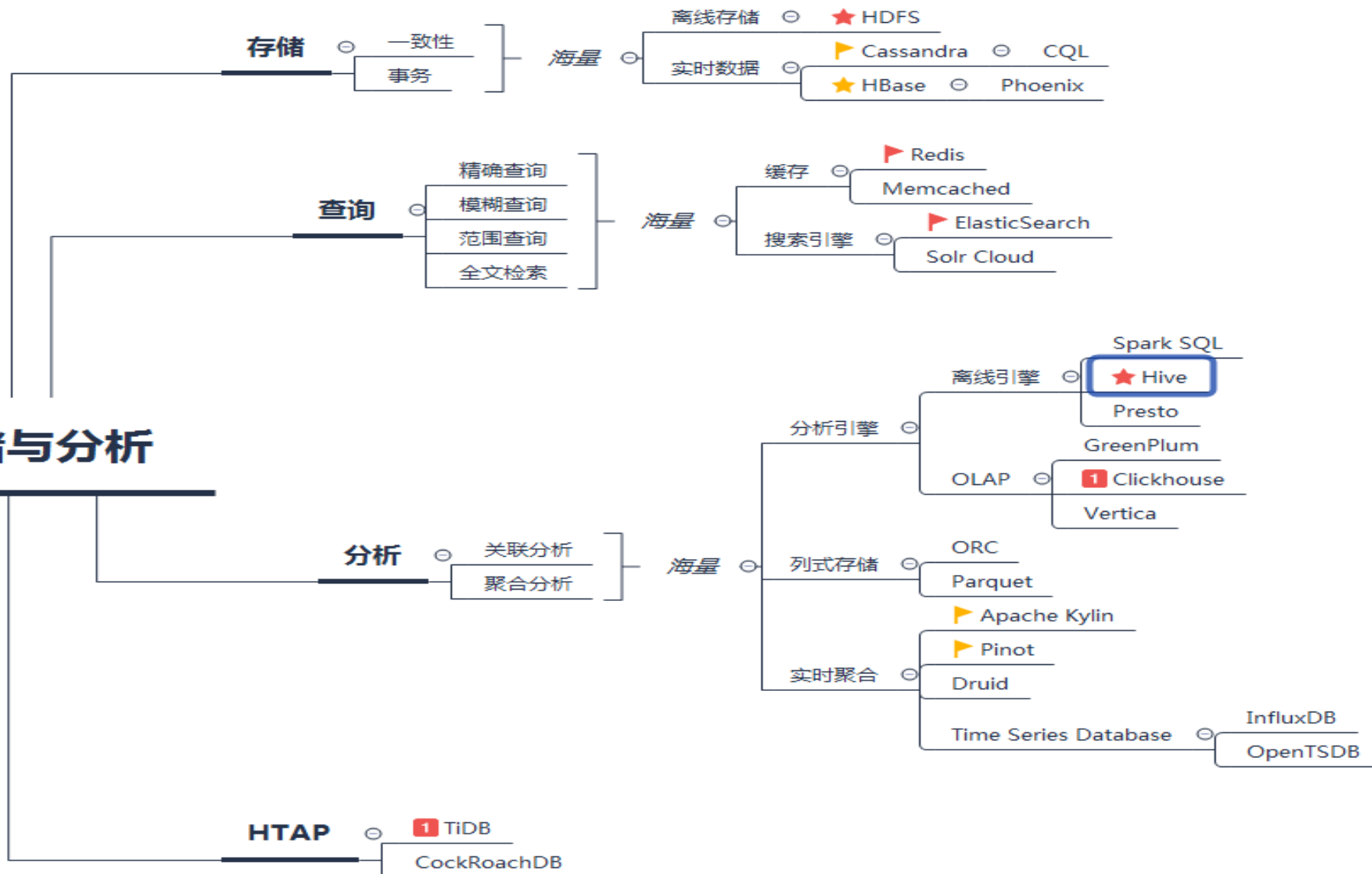
PRESENTED BY XU PENG

# ELASTICSEARCH在数据湖中的地位



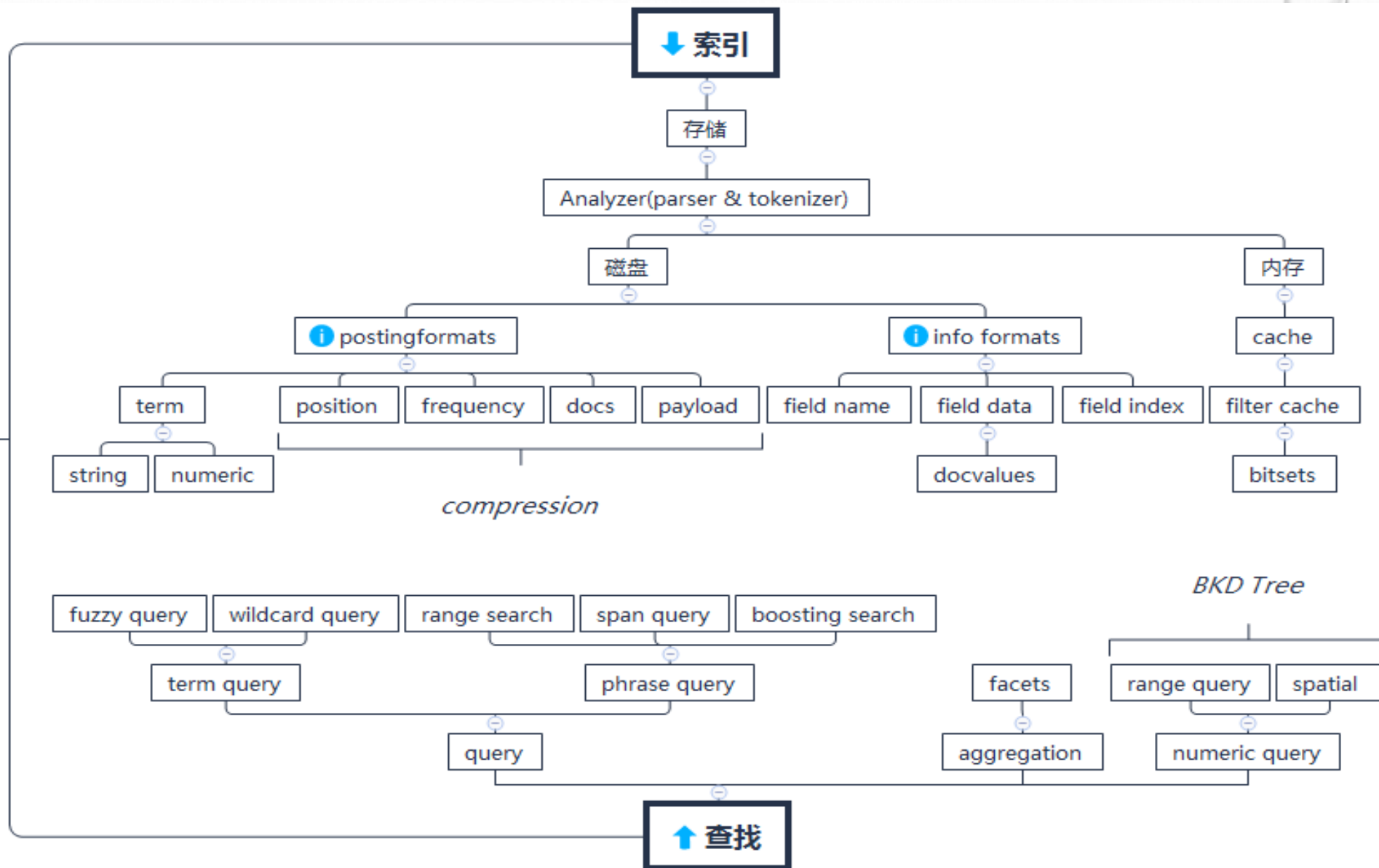
# ELASTICSEARCH的功能

## 数据存储与分析



# LUCENE核心概念

→ 文档

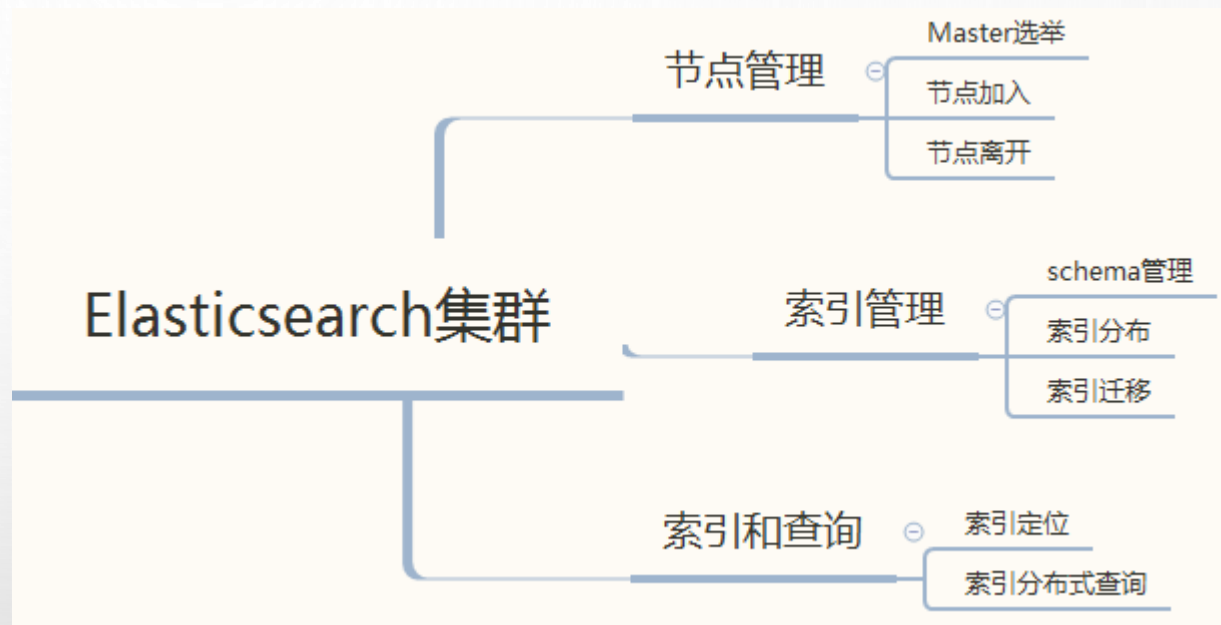


↑ 查找

## PART II ELASTICSEARCH集群

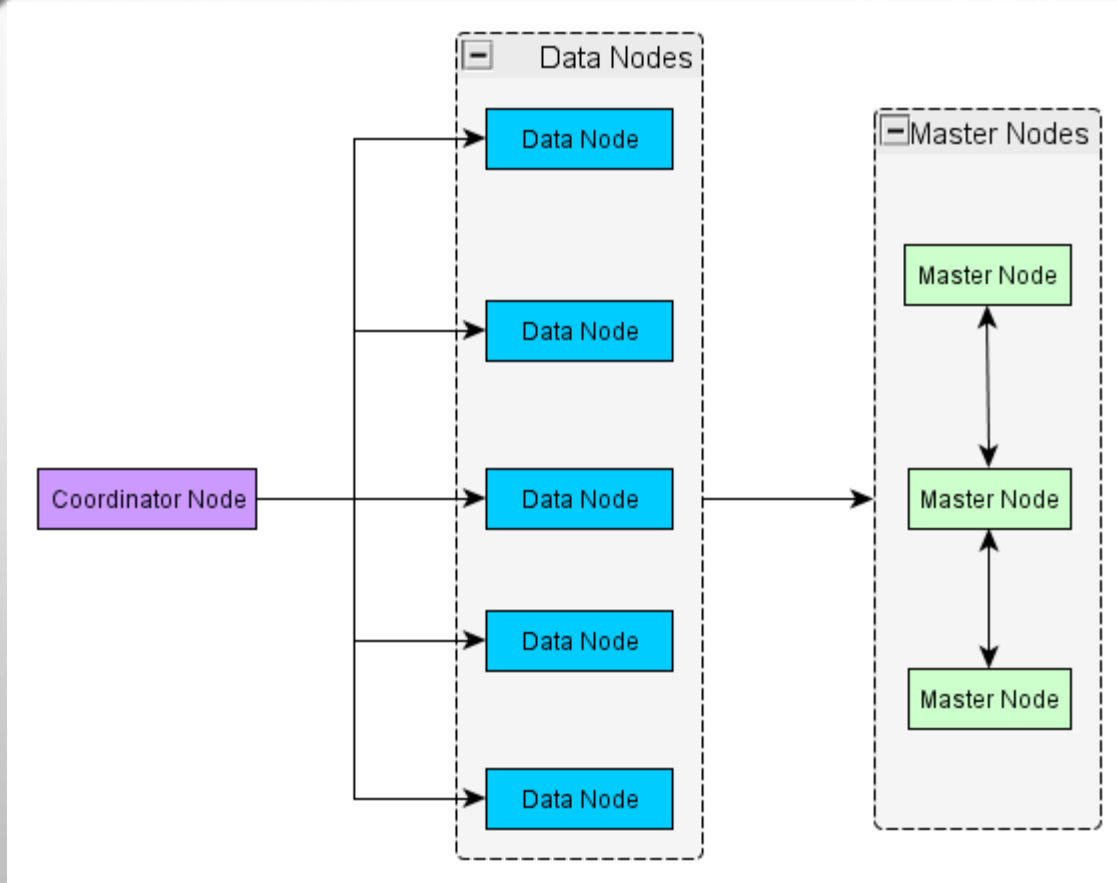
# ELASTICSEARCH – 分布式LUCENE索引

- 集群管理，节点层面
- 分布式索引管理
  1. Schema管理
  2. 索引分布
  3. 索引迁移
- 索引查询
  - query-and-fetch



# ELASTICSEARCH集群节点组成

- 节点类型
  - Master
  - Data
  - Coordinator
  - Ingest



# 节点发现机制

- ZEN DISCOVERY
- MASTER节点包含CLUSTER状态信息
  1. 节点的加入或离开
  2. INDEX的创建、删除、打开、关闭
  3. SHARDS的分配和路由信息
  4. SCHEMA改变
- `CURL -XGET LOCALHOST:9200/_CLUSTER/STATE`

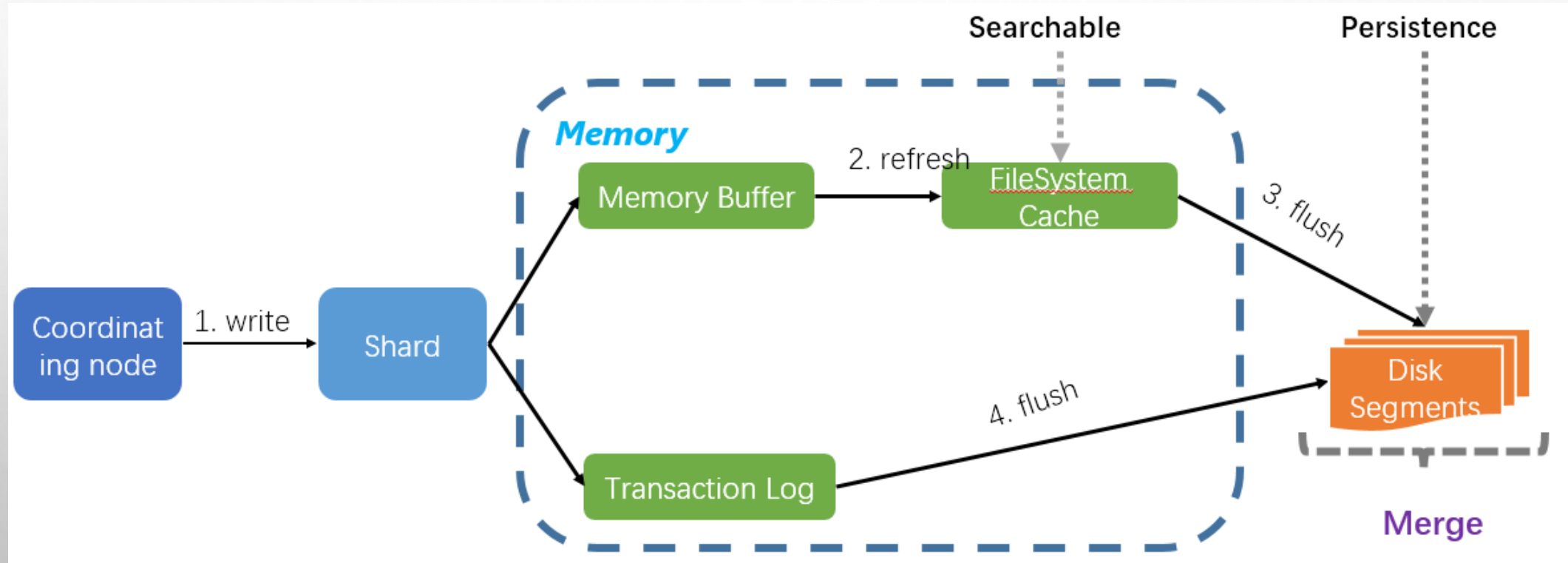


# 最精简的ES配置

*#配置文件/etc/elasticsearch/elasticsearch.yml*

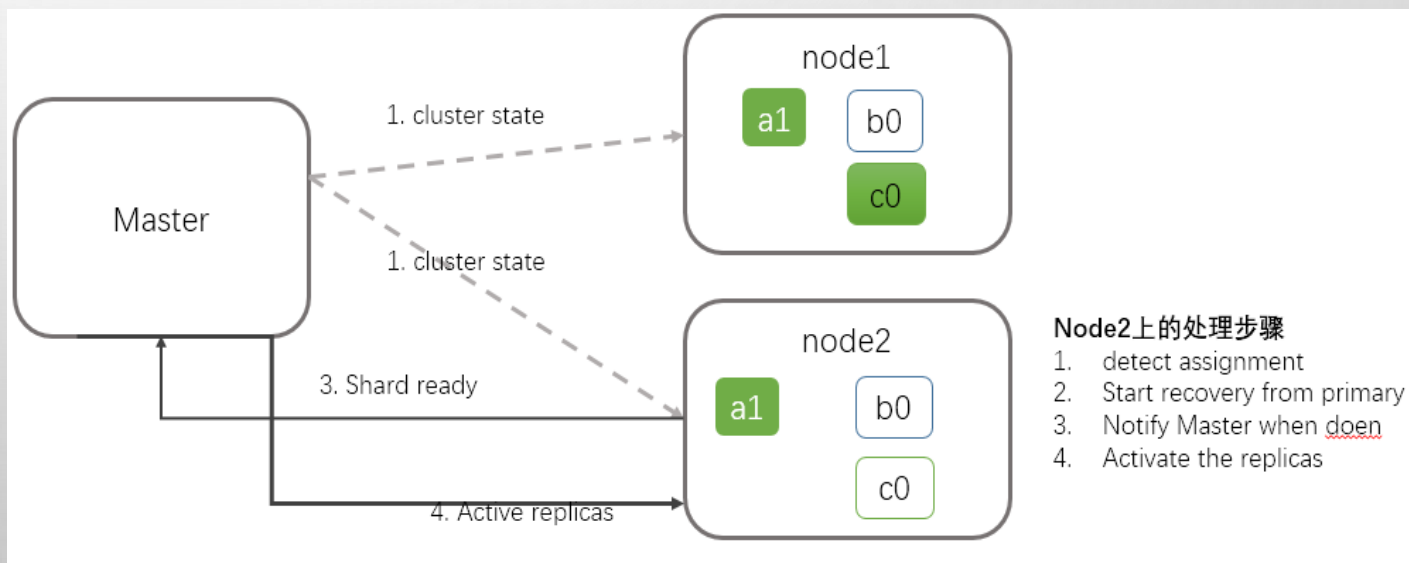
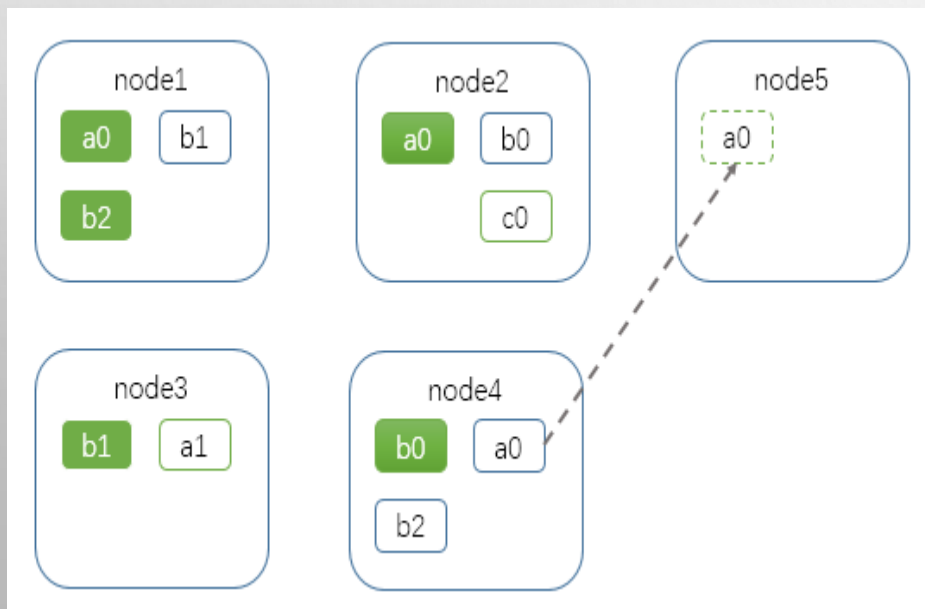
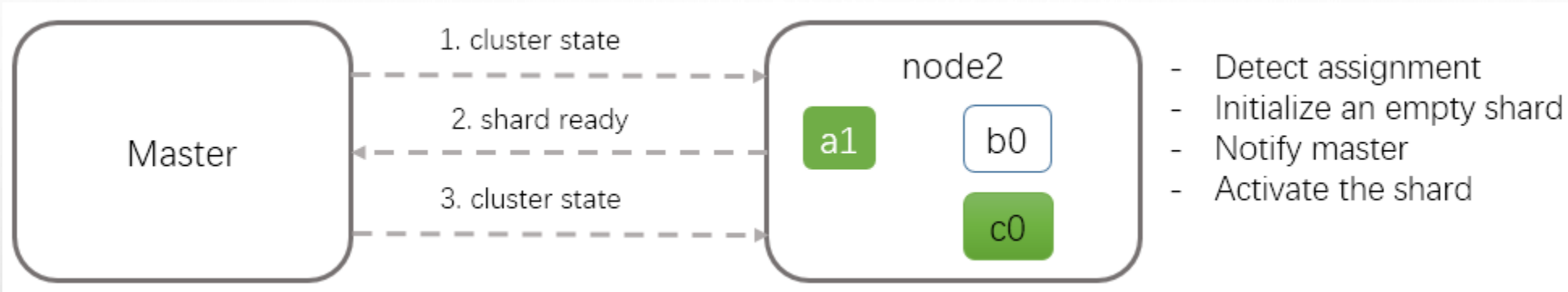
```
cluster.name: es_demo_cluster
node.name: es_demo_node_1
node.master: true
node.ingest: false
node.data: true
bootstrap.memory_lock: true
bootstrap.system_call_filter: false
network.host: 192.168.56.101
discovery.zen.ping.unicast.hosts: ["192.168.56.101","192.168.56.102","192.168.56.103"]
discovery.zen.minimum_master_nodes: 2
#xpack配置
xpack.security.enabled: false
xpack.monitoring.enabled: true
xpack.graph.enabled: false
xpack.watcher.enabled: false
xpack.monitoring.exporters.my_remote.type: http
xpack.monitoring.exporters.my_remote.host: ["http://localhost:9200"]
```

# 索引写入过程分析



# 分片管理

## 分片初始化



# INDEX SETTINGS

```
{  
  "settings": {  
    "index": {  
      "routing": {  
        "allocation": {  
          "total_shards_per_node": "3"  
        }  
      },  
      "refresh_interval": "120s",  
      "number_of_shards": "18",  
      "translog": {  
        "flush_threshold_size": "2g",  
        "durability": "async",  
        "sync_interval": "15s"  
      },  
      "merge": {  
        "scheduler": {  
          "max_thread_count": "1"  
        }  
      },  
      "number_of_replicas": "0",  
    }  
  }  
}
```

## - checklist

1. 避免同一个index的所有shard落入同一个数据节点
2. 根据业务场景动态调整refresh\_interval
3. 调整flush\_threshold\_size大小
4. 每一个shard占用的磁盘空间控制在10GB~15GB
5. 同一个节点所管理的分片数不超过600个，20 shard/per gb heap
6. 注意索引写入过程中Throttling的次数
7. 根据历史统计信息，动态调整索引配置，维护一个健康稳定的集群
8. 不要把Elasticsearch做为核心关键数据的主要存储

# MAPPINGS – 定义 SCHEMA

```
"mappings": {  
  "logs": {  
    "properties": {  
      "builtinTimestamp": {  
        "type": "date",  
        "format": "YYYY-MM-dd HH:mm:ss.SSS"  
      },  
      "exception": {  
        "type": "keyword",  
        "ignore_above": 10915  
      },  
      "request": {  
        "type": "text",  
        "index": false  
      },  
      "response": {  
        "type": "keyword"  
      },  
      "timestamp": {  
        "type": "date"  
      },  
    }  
  }  
}
```

- `_source`
  - `enabled`
  - 重建索引时有用，最好不要禁止
- `properties`
  - `Text`与`keyword`的区别
  - `index`
  - `doc_values`

Notes: 时间类型的格式

"yyyy-MM-dd HH:mm:ss | | yyyy-MM-dd | | epoch\_millis"

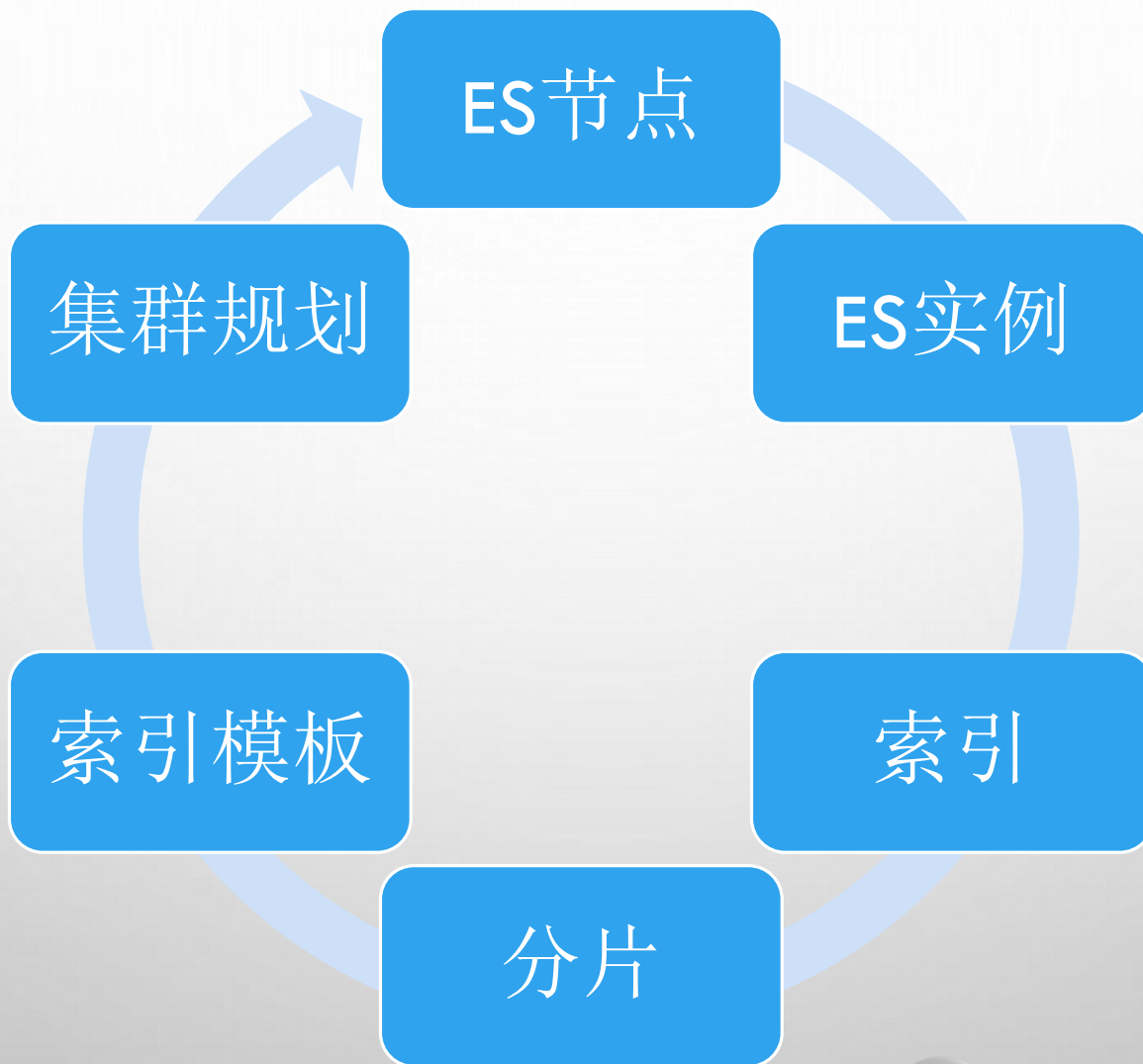
# 集群索引恢复

```
"transient": {  
  "cluster": {  
    "routing": {  
      "allocation": {  
        "node_initial primaries recoveries": "25",  
        "balance": {  
          "index": "16.0f",  
          "threshold": "1.0f",  
          "shard": "0.02f"  
        },  
        "enable": "all",  
        "cluster_concurrent_rebalance": "120",  
        "node_concurrent_recoveries": "4",  
        "exclude": {  
          "_name": "",  
          "_ip": "192.168.0.101, 192.168.0.102"  
        }  
      }  
    }  
  }  
}
```

*值越大，意味着某一个索引的shard在集群中分布越均匀*

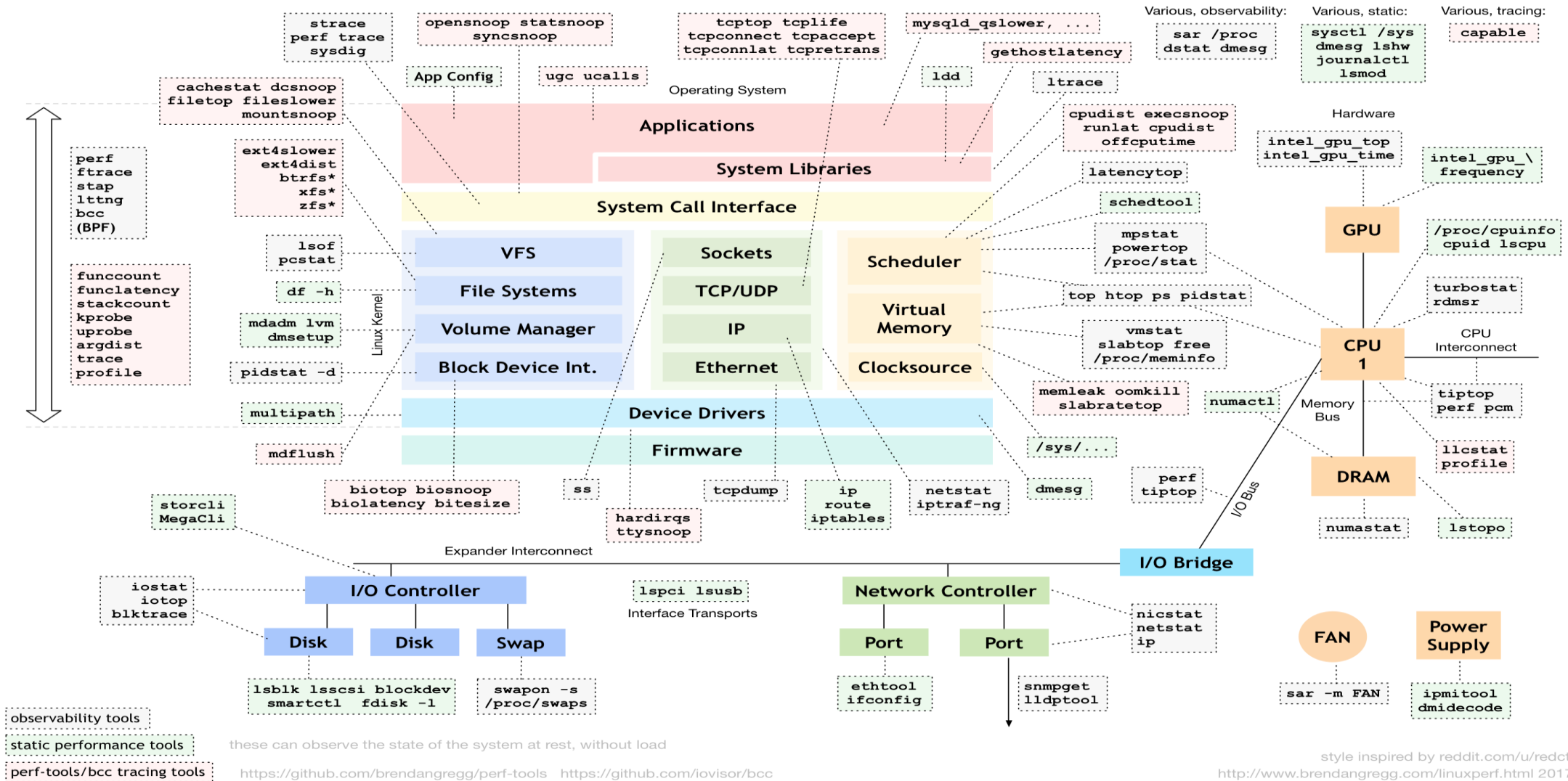
*值越大，集群中某一个节点上的shard数目越均匀*

# ES监控要点



# OS监控工具一览

Linux Performance Tools





# 集群配置和监控

## - OS参数配置

### - 内存

`vm.dirty_ratio`

`vm.dirty_background_ratio`

### - I/O scheduler

```
echo 'cfq' > /sys/block/sd$/queue/scheduler
```

```
blockdev --setra 1024 /dev/sd$
```

Ssd硬盘，推荐使用noop

### - 关闭Transparent Huge Pages

```
echo 'never' > /sys/kernel/mm/transparent_hugepage/defrag
```

```
echo 'never' > /sys/kernel/mm/transparent_hugepage/enabled
```

### - 禁用numa

```
sudo sysctl -w vm.zone_reclaim_mode=0
```

## - JVM设置

1. 内存不超过32GB

2. 避免调整thread\_pool

## - 监控工具

1. htop/atop

2. sar

3. perf

# CAT API

## 集群状态

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

## 索引信息

GET \_cat/indices  
GET \_cat/segments  
GET \_stats

## 查询具体参数

*GET \_cat/nodes?help*

## 节点状态

GET \_nodes/stats 统计信息  
GET \_nodes/node\_name/stats 具体某一个节点的信息  
GET \_nodes/ 基础配置信息











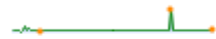







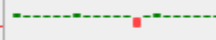
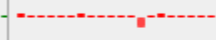


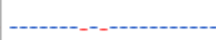





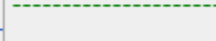
















## shard信息

GET \_cat/shards

# 监控方案 节点信息

Load	CPU	Indexing	Search	I/O Ops	Old GC	Young ...	Refresh	Merges	Flush	Memory Usage	
										Segme...	Query ...
107.86	4									1.28GB	4.13MB
107.86	0									849.68MB	
107.86	4									1.31GB	3.14MB
37.29	0									1.30GB	5.36MB
37.29	0									1.32GB	3.19MB
32.5	0									835.03MB	
31.82	100									1.35GB	7.09MB
31.82	3									1.28GB	2.72MB
29.31	2									1.33GB	4.23MB
29.31	5									1.31GB	4.60MB
25.91	6									1.21GB	2.44MB
										1.24GB	6.12MB

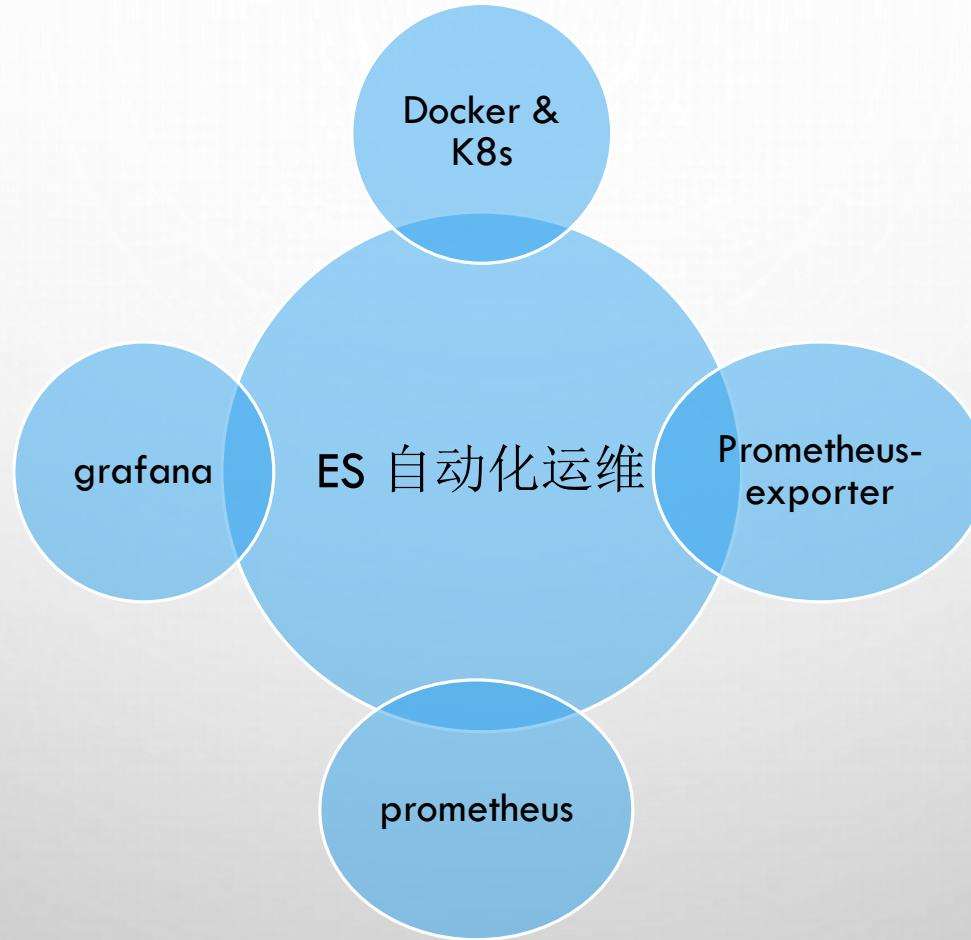
# 监控方案 索引统计信息

Query Rate ▼	Query ▲	Segments		Translog		store ▲	Throttle ▲	Segments ▲	Writer Me... ▲
		refresh ▲	merge ▲	flush ▲	flush peri... ▲				
 16.73	30613					1598.19GB	2889690	987.10MB	5.23GB
 14.40	6942					8.92GB	0	16.95MB	201.26MB
 11.11	6928					1269.57GB	0	153.03MB	356.63MB
 7.80	2464					1404.65GB	0	59.43MB	96.61MB
 7.11	2781					910.95GB	1795	159.84MB	805.66MB
 7.06	5697					1.85GB	0	6.98MB	147.22MB
 5.89	4888					797.21GB	0	80.23MB	90.40MB
 4.53	2347					340.67GB	0	125.72MB	764.79MB
 4.53	1751					141.55GB	0	123.43MB	592.07MB

# 集群状态 日志文件

```
[2019-04-12T10:14:28,867][INFO ][o.e.c.s.ClusterApplierService] [ctrip_flight_v5_demo_183_98]
removed
{{ctrip_demo_v6_data_0_185}{MNhNRMSPQpKbv2ukGewKzw}{1NU_GBxCQwS7Btxg63osvA}{192.168.0.185}{192.168
.0.185:9301}{ml.machine_memory=134610358272, ml.max_open_jobs=20, xpack.installed=true,
box_type=hot, ml.enabled=true}}, reason: apply cluster state (from master [master
{ctrip_flight_v6_demo_183_98}
```

# 自动化/智能化运维



The background is a light gray gradient with several realistic water droplets of various sizes scattered in the corners. The droplets have highlights and shadows, giving them a three-dimensional appearance. The text is centered on the page.

Q & A

THANK YOU



elastic  
中文社区

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