

# ES内部分享

---

## 1. ES简介

ElasticSearch是Elastic公司开发的开源分布式搜索引擎

- 开源
- 分布式
- 全文检索
- OLAP(结合kibana使用)
- Resful API
- NoSQL database

### 1.1 和Lucene的关系

```
{
  "name": "uf_-1wJ",
  "cluster_name": "UT",
  "cluster_uuid": "xvGp84DyQV0SxLXP43oDpA",
  "version": {
    "number": "6.2.4",
    "build_hash": "ccec39f",
    "build_date": "2018-04-12T20:37:28.497551Z",
    "build_snapshot": false,
    "lucene_version": "7.2.1",
    "minimum_wire_compatibility_version": "5.6.0",
    "minimum_index_compatibility_version": "5.0.0"
  },
  "tagline": "You Know, for Search"
}
```

简单而言，ES是Lucene的分布式版本，当然扩充了接口和在线分析功能

### 1.2 基本概念

#### 1.2.1 索引层面

- index
- type
- doc
- term

#### 1.2.2 倒排索引

正常顺序 doc -> term

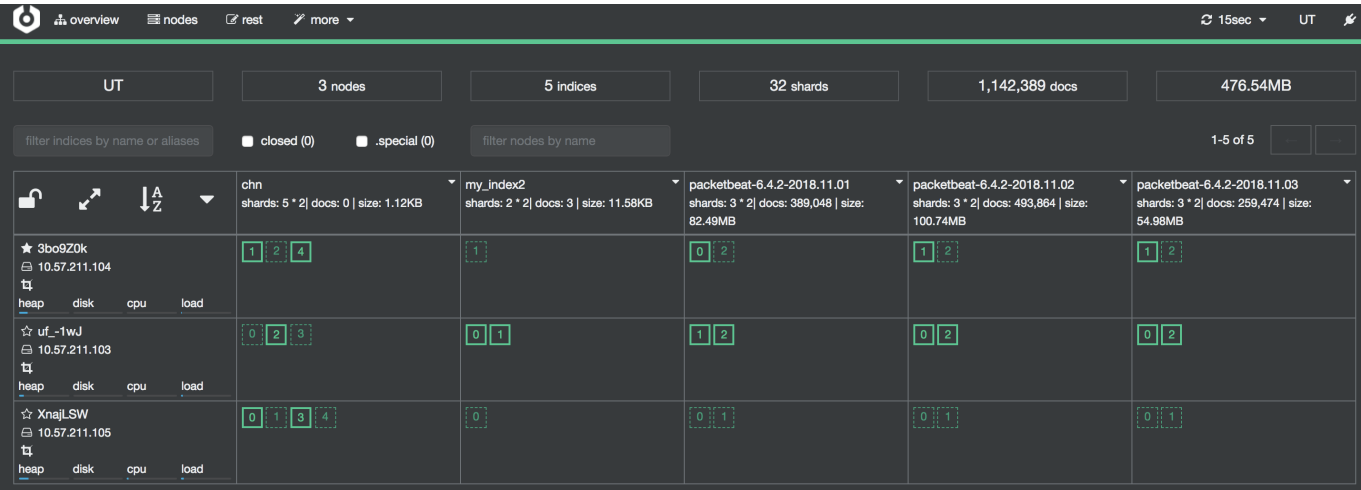
term -> term

1.2.2 集群层面

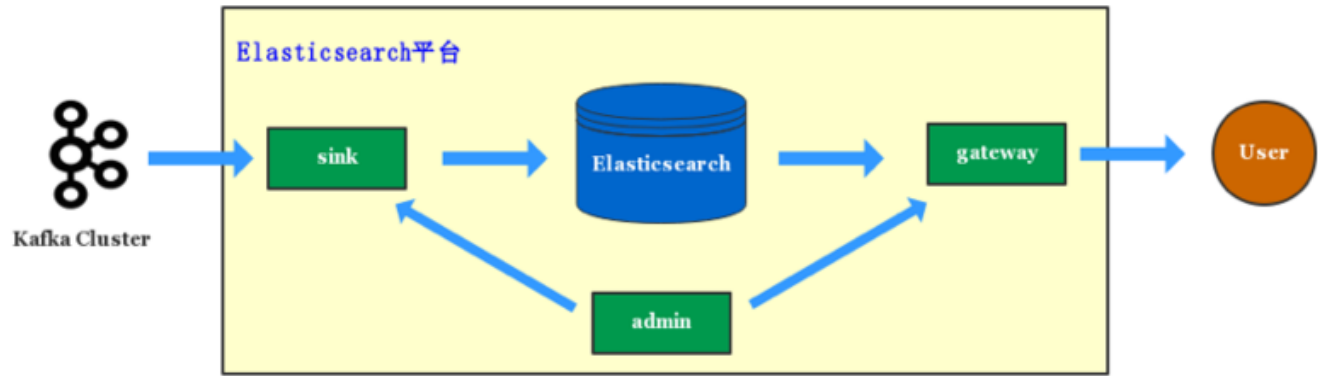
Cluster Node Shard Segment

多种角色

- 1. Master-eligible node
- 2) Data node
- 2. Ingest node
- 3. Tribe node
- 4. coordinating node



1.3 通常的架构

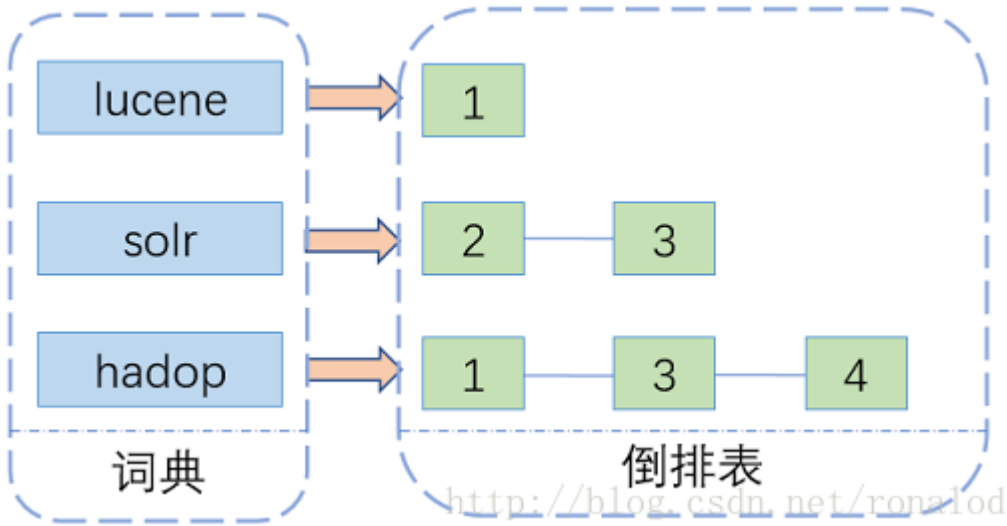


2. 建立索引&查询

2.1 Lucene底层结构

2.1.1 字符串

数据结构: SSTable



2.1.2 数值型

数据结构: Block K-D tree

- 1. 时间类型字段 2) 数值型字段的存储(float、double、integer、long)
- 2. 多维向量

2.2 查询语法

Lucene查询语法

```
age: [10,20] AND name: 李四
```

ES的查询语法

```
{
  "from": 0,
  "size": 200,
  "query": {
    "bool": {
      "must": [{
        "range": {
          "age": {
            "from": 10,
            "to": 20
          }
        }
      }],
      "filter": [
        {
          "match": {
            "name": "李四"
          }
        }
      ]
    }
  }
}
```

```
}  
}  
}
```

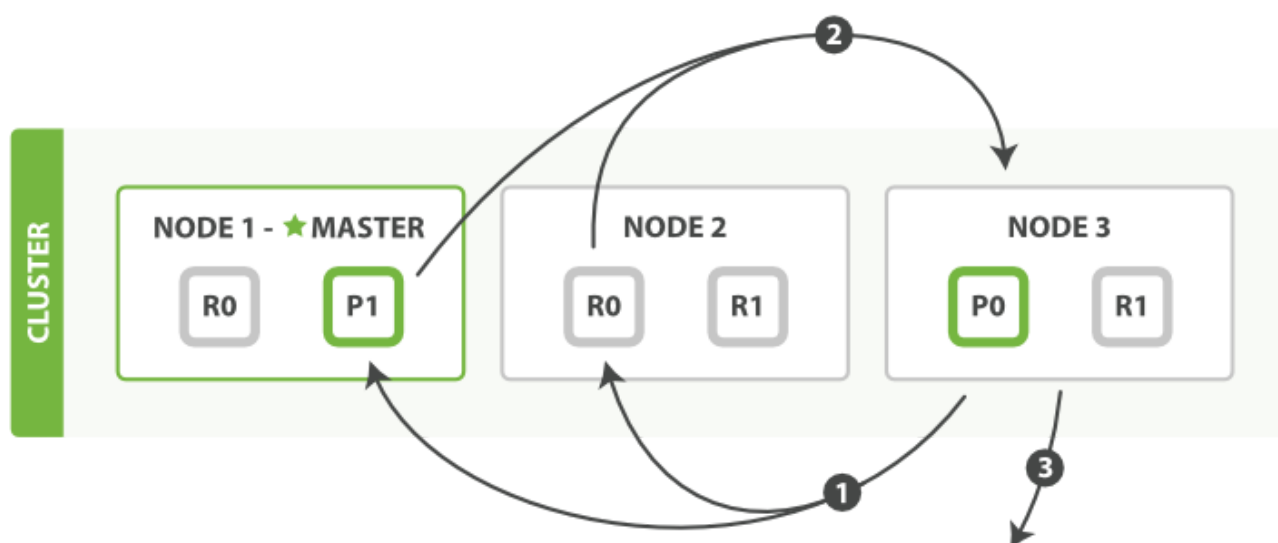
```
{  
  "query": {  
    "query_string" : {  
      "query" : "age:[10,20] AND name:李四"  
    }  
  }  
}
```

```
{  
  "query": {  
    "wildcard": {  
      "name": {  
        "value": "王*"  
      }  
    }  
  }  
}
```

term/match 的区别

## 2.3 查询过程

query and fetch



2.3 关于分词器的作用(Analyzer)

## 3. SQL&Luke

3.1 SQL

Elasticsearch-sqlhttp://localhost:9200Help

SQL Query

```
1 SELECT COUNT(*) AS count,
2 AVG(balance) AS average_balance,
3 SUM(balance) AS sum_balance
4 FROM elasticsearch-sql test index
5 WHERE (age >= 27 AND age<=30) OR city LIKE 'a%'
6 GROUP BY gender, age
```

Search

Results

gender	age	count	sum_balance	average_balance
f	28	31	839381	27076.8064516129
f	30	25	632904	25316.16
f	27	21	453996	21618.85714285714
f	29	12	343217	28601.416666666668
f	26	3	14433	4811
f	36	2	62717	31358.5
f	20	1	34473	34473

3.2 Luke

Luke: Lucene Toolbox Project - v7.6.0

File Tools Help

Overview Documents Search Analysis Commits Logs

Index Path: .../solr-7.6.0/example/techproducts/solr/techproducts/data/index

Number of Fields: 23

Number of Documents: 32

Number of Terms: 950

Has deletions? / Optimized?: No / Yes

Index Version: 6

Index Format: Lucene 7.4 or later

Directory implementation: org.apache.lucene.store.MMapDirectory

Currently opened commit point: segments\_2 (generation=2, segs=1)

Current commit user data: {commitCommandVer=1619925510184239104, commitTimeMSec=1544881353554}

Select a field from the list below, and press button to view top terms in the field.

Available fields and term counts per field:

Name	Term count	%
text	399	42.00 %
features	293	30.84 %
name	111	11.68 %
id	32	3.37 %
manu	31	3.26 %
manu_e...	16	1.68 %
cat	16	1.68 %
manu_id_s	14	1.47 %
compNa...	11	1.16 %
address_s	10	1.05 %
includes	9	0.95 %
price...	4	0.42 %
payloads	2	0.21 %
inStock	2	0.21 %

Selected field:

text

Show top terms >

Num of terms:

50

Top ranking terms: (Double-click for more options.)

Rank	Freq	Text
1	14	electronics
2	8	inc
3	8	and
4	5	x
5	5	usb
6	5	one
7	5	notes
8	5	memory
9	5	lcd
10	5	2.0
11	4	with
12	4	to
13	4	tft
14	4	sdram

4. 参考资料

4.1 Elasticsearch Reference

4.2 中文官方社区

4.3 ElasticSearch源码解析与优化实践



4.4 elasticsearch 权威指南

已经过时

4.5 Lucene实战

