马士兵教育-Elasticsearch

——高手进阶篇

**深入解析relevance score原理以及排序规则优化**

1. shard local idf和global idf（多shard下评分不准确问题解析）
2. multi\_match多字段搜索
   1. best\_fields、most\_fields和cross\_fields策略
      1. best\_fields：对于同一个query，单个field匹配更多的term，则优先排序。
      2. most\_fields：如果一次请求中，对于同一个doc，匹配到某个term的field越多，则越优先排序。
      3. cross\_fields：

#吴必须在 姓 中出现或者 名 中出现 或者

#磊必须在 姓 中出现或者 名 中出现。

* 1. dix\_max查询（Disjunction Max Query）：将任何与任一查询匹配的文档作为结果返回，但只将最佳匹配的评分作为查询的评分结果返回
  2. tie\_breaker：取值范围 [0,1]，其中 0 代表使用 dis\_max 最佳匹配语句的普通逻辑，1表示所有匹配语句同等重要。最佳的精确值需要根据数据与查询调试得出，但是合理值应该与零接近（处于 0.1 - 0.4 之间），这样就不会颠覆 dis\_max 最佳匹配性质的根本。

1. function score query：必须定义​​一个查询和一个或多个函数，自定义函数会为查询返回的每个文档计算一个新分数
   1. grammar：见代码
   2. weight：设置权重
   3. field\_value\_factor：将某个字段的值进行计算得出分数。
      1. field：要计算的字段
      2. modifier：以何种运算方式计算，接受以下枚举
         1. none：不处理
         2. log：计算对数
         3. log1p：先将字段值 +1，再计算对数
         4. log2p：先将字段值 +2，再计算对数
         5. ln：计算自然对数
         6. ln1p：先将字段值 +1，再计算自然对数
         7. ln2p：先将字段值 +2，再计算自然对数
         8. square：计算平方
         9. sqrt：计算平方根
         10. reciprocal：计算倒数
      3. factor：当前分数计算，对整个结果产生的权重比
   4. random\_score：随机得到 0 到 1 分数
   5. decay Function：同样以某个字段的值为标准，距离某个值越近得分越高
   6. script\_score：通过自定义脚本计算分值
   7. boost\_mode：指定计算后的分数与原始的\_score如何合并，有以下选项：
      1. multiply：查询分数和函数分数相乘
      2. sum：查询分数和函数分数相加
      3. avg：取平均值
      4. replace：替换原始分数
      5. min：取查询分数和函数分数的最小值
      6. max：取查询分数和函数分数的最大值
   8. max\_boost：分数上限