



ElasticSearch

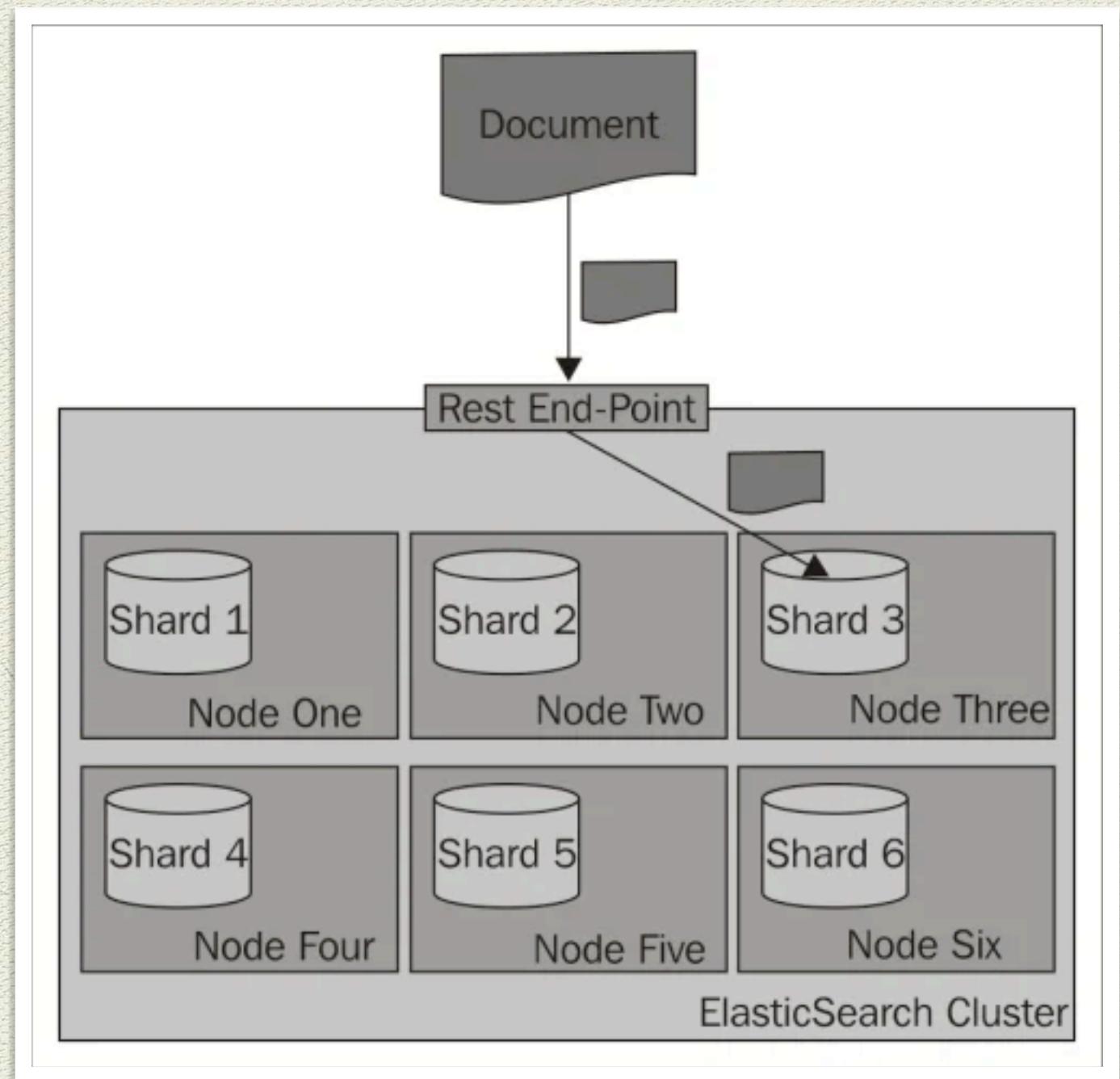
完全入门教程

09. Shard大小规划

波波微课

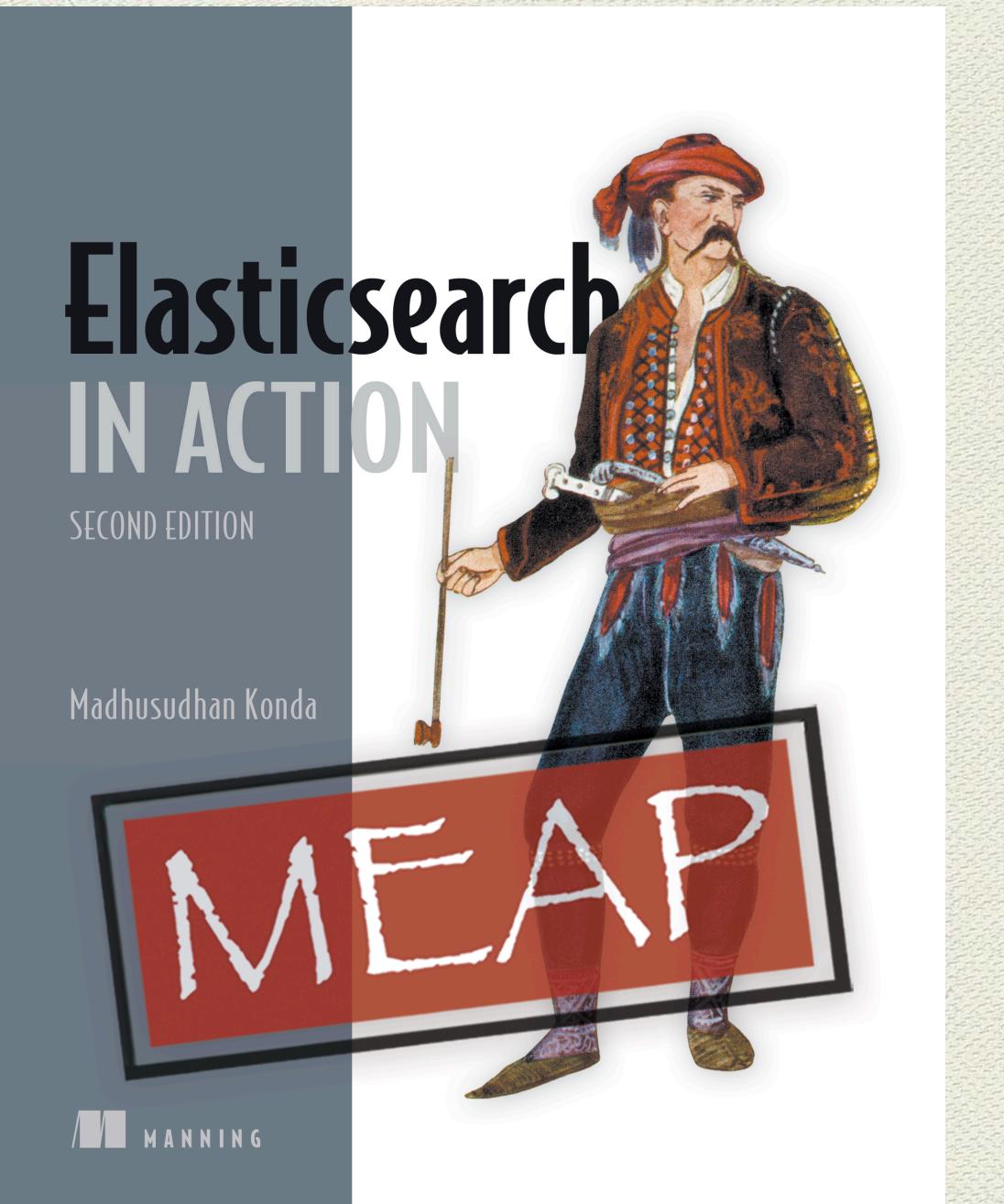
活跃Index上的Shard数量无法简单修改

- ◆ 路由算法(索引和查询)
 - ◆ $\text{shard_number} = \text{hash}(\text{document_id}) \% \text{number_of_primary_shards}$
 - ◆ 依赖于shard数量，运行时修改会导致文档错误
- ◆ 修改Shard数量需要使用**re-indexing**
- ◆ 运行时可以修改replica数量
- ◆ Index settings API



选择合适的Shard大小

- ◆ 取决于组织的当前和未来的数据需求，可能需要试错
- ◆ 实践
 - ◆ 业界推荐单个shard <=50GB
 - ◆ Github实践128个shards，每个120GB
 - ◆ 作者推荐25G ~ 40GB(考虑堆内存需求)
 - ◆ 500GB数据 -> 10 ~ 20个Shards
- ◆ 堆内存需求
 - ◆ 推荐每GB堆内存支持最多20个Shards
 - ◆ 设置jvm.options(Xms / Xmx)，ES缺省1GB



存储需求规划

- ◆ 3 Shards + 3 Replicas/Shard, 每个Shard 50GB
- ◆ Shards存储需求 $3 \times 50 \text{ GB} = 150\text{GB}$
- ◆ Replicas存储需求 $3 \times 3 \times 50\text{GB} = 450\text{GB}$
- ◆ 还需考虑：系统存储 + 堆内存