



Introdução ao Elasticsearch na prática

Rafael Mota



<https://github.com/faelmota/minicurso-elasticsearch>



Um pouco sobre mim

- Formado em Sistemas de Informação, 2012 (FEPI, Itajubá)
- Pós graduado em Gestão de Projetos, PMI, 2014 (FAI, Santa Rita do Sapucaí)
- Desenvolvedor na IWS desde 2012



O que é o Elasticsearch (ES)

- Mecanismo de busca e análise de dados
- Código aberto
 - <https://github.com/elastic/elasticsearch>
- Faz parte do Elastic Stack (ELK)
- API simples baseada em REST



Integrando com outras aplicações

- Pode ser facilmente integrado com qualquer aplicação
- Algumas bibliotecas:
 - PHP (composer require elasticsearch/elasticsearch)
 - Node (npm install @elastic/elasticsearch)
 - Python (python -m pip install elasticsearch)
 - Java (dependencies {
 - implementation 'co.elastic.clients:elasticsearch-java:8.13.2'
 - implementation 'com.fasterxml.jackson.core:jackson-databind:2.17.0'
 - })



Terminologia

Elasticsearch	Banco de dados relacional (BDR)
Índices	Tabelas
Documento (JSON)	Linha
Campo	Coluna
Mapeamento	Estrutura da tabela



Tipos de dados

- String
- Integer
- Float
- Date
- Boolean
- Object
- Array
- Binary



Exemplos de uso

- Em aplicações
 - Campos de pesquisas
 - Avaliações
 - Últimos vistos/Mais vendidos
 - Filtros
- Logs
 - Arquivos de logs (Nginx, apache)
- Gerenciamento de desempenho de aplicativos (APM)
 - CPU/Memória /Disco
- Relatórios
 - Vendas, compras, etc.



Índice invertido

- Estrutura de dados que permite a rápida recuperação de documentos com base em termos de pesquisa
- Cada termo é mapeado para os documentos que o contêm
- O ES consulta o índice invertido para encontrar rapidamente os documentos relevantes



“Minicurso - Introdução ao Elasticsearch na prática”

[“minicurso”, “introdução”, “elasticsearch”, “prática”]

Termo	Documentos
minicurso	1
introdução	1,3
elasticsearch	1,2,4,5,6
prática	1,3



Sharding

- Sharding é a maneira de dividir índices em fatias menores (shards)
 - É realizado a nível de índice
- Permite escalonamento horizontal
 - Mais documentos
 - Mais índices
 - Melhora o desempenho

Node 1 (Master)

P1

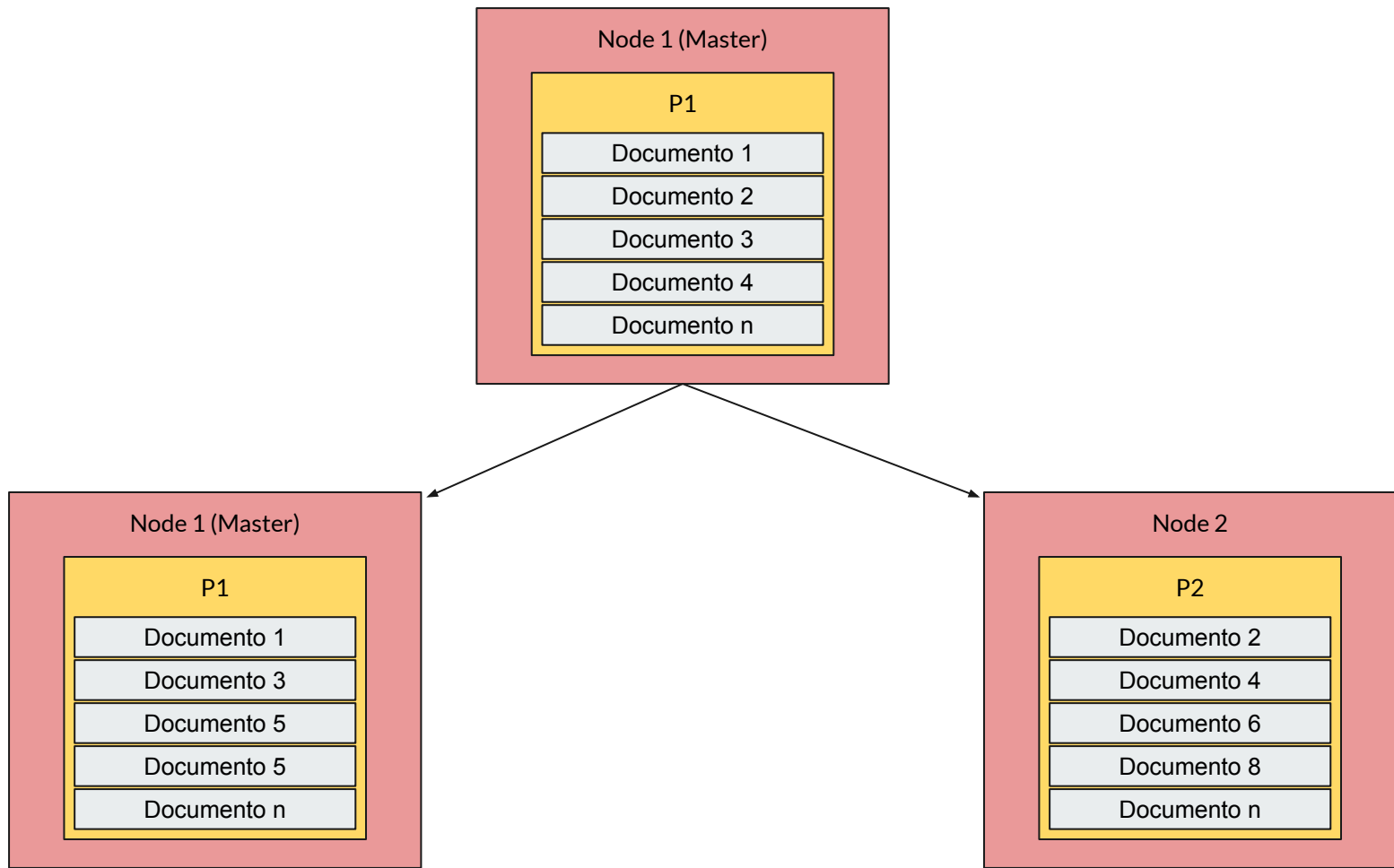
Documento 1

Documento 2

Documento 3

Documento 4

Documento n





Réplicas

- Réplicas são cópias dos shards
 - Garantem disponibilidade
 - Tolerância a falhas
 - Redundância
 - Distribuição de cargas

Node 1 (Master)

P1

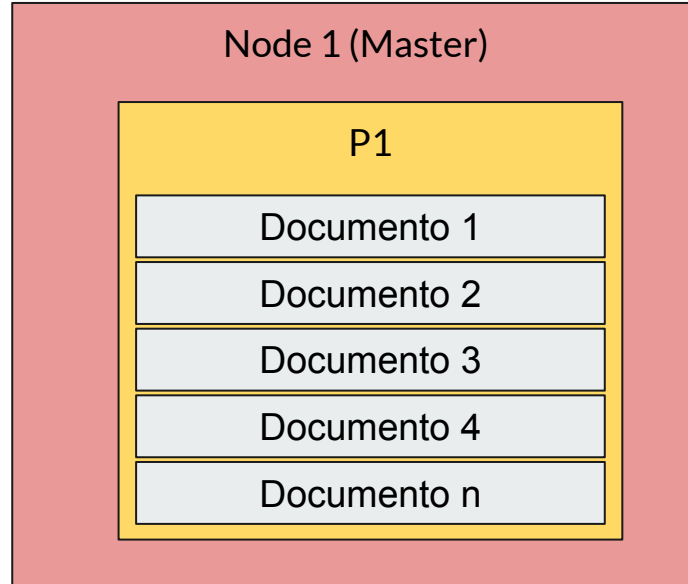
Documento 1

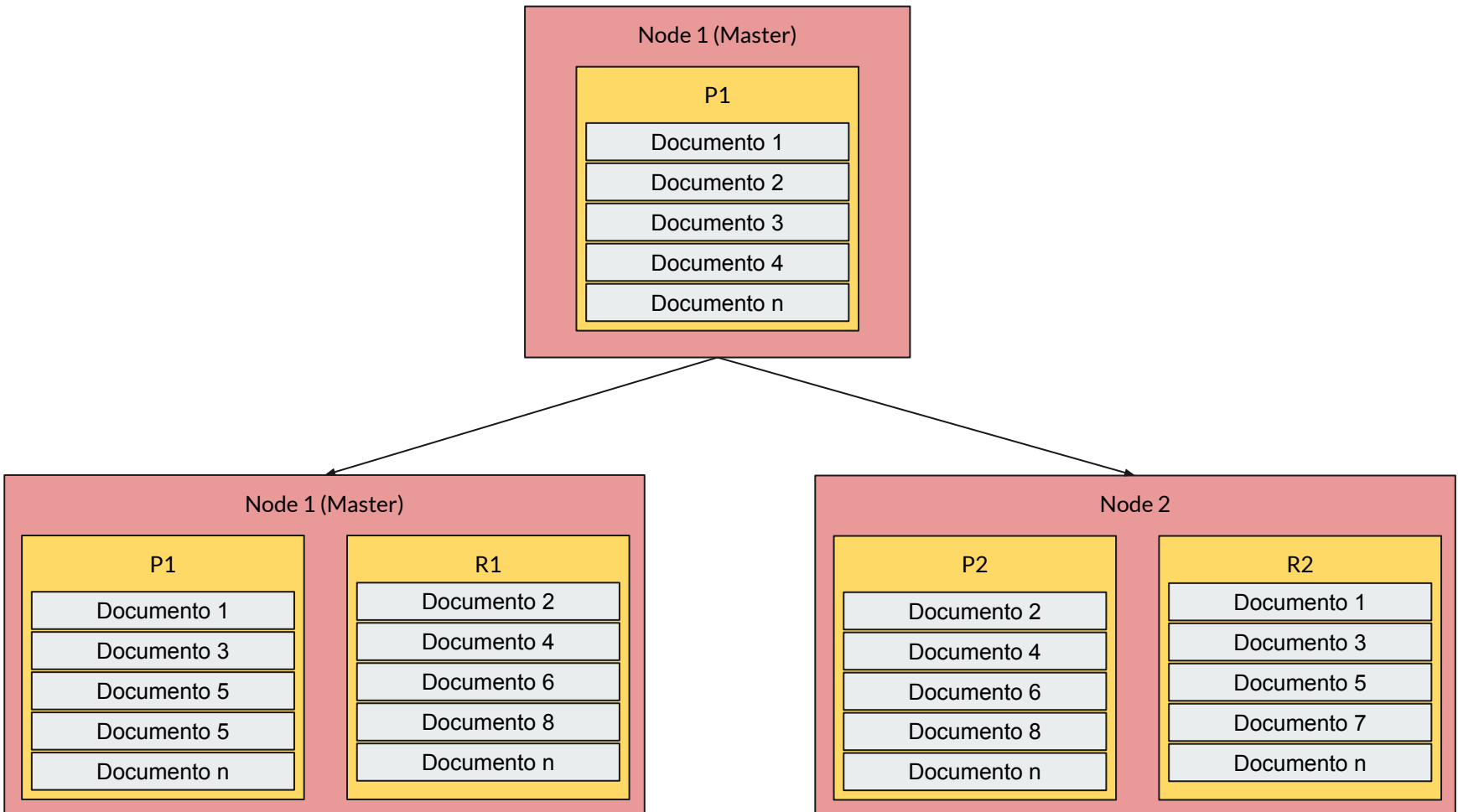
Documento 2

Documento 3

Documento 4

Documento n





Mão na massa
