

Elastic Search

Seminar Presentation

Mohammed Nisham

July 28, 2016

College of Engineering, Trivandrum

Table of Contents

1. Introduction
2. Features
3. Search
4. Performance Analysis
5. Conclusion

Introduction

What is elastic search

- Built on top of Apache Lucene [1]
- Built in java, Uses RESTful APIs
- Users - Facebook, Github, Wikipedia
- Ranked first in Search engine Databases
- Full text search

- Shay Banon created Compass library in Java
- Not Scalable
- Rebuilt in distributed approach using RESTful API

Features

- Documents as JSON object
- Index
- Type - mapping
- Id
- Dynamic mapping

- Complete Lucene search engine
- Distributed document store
- Immutability
- Near real timing by In-memory Buffer
- Crash recovery with Translog

Sharding

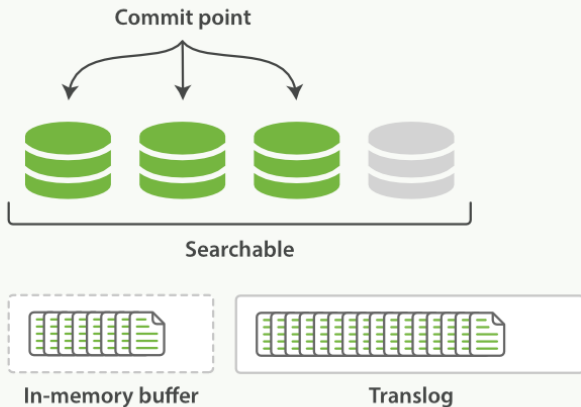


Figure 1: Inside a shard

Clustering

- Node - instance of ES server
- Primary and Replica shards
- Failure recovery
- Horizontal scaling
- Cluster master
- Completely autonomous

Clustering

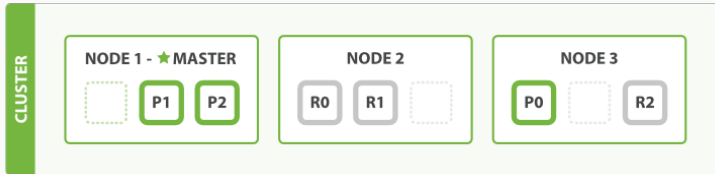


Figure 2: A sample cluster

Clustering

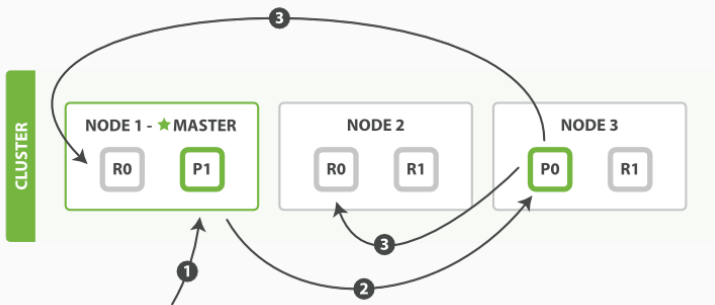


Figure 3: Request routing in a cluster

- Buckets
- Metrics
- Usable with filters and queries

- Optimistic concurrency control
- Operations are asynchronous and concurrent
- Metafield version

Search

- Score Calculation
- Term Frequency
- Inverted Document frequency
- Field length norm

- Query phase
- Fetch phase

Distributed search

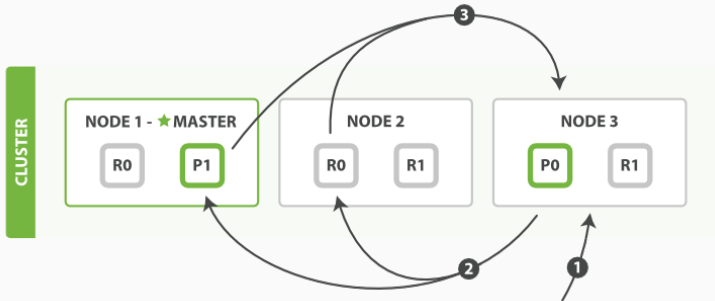


Figure 4: Query phase in distributed search

Distributed search

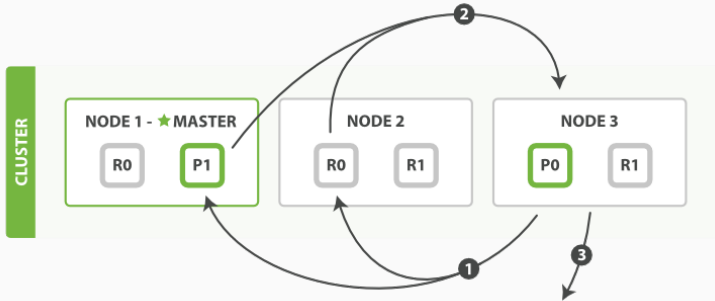


Figure 5: Fetch phase in distributed search

- Query DSL
- Query and Filter
- Combinations

- Inverted Index
- Multi word search
- Multi field search
- Phrase search
- Search as you type, Edge n-grams
- Fuzzy search

Term	Doc1	Doc2	Doc3
Brown	X		X
quick	X	X	X

Figure 6: An inverted index

Performance Analysis

Performance Analysis

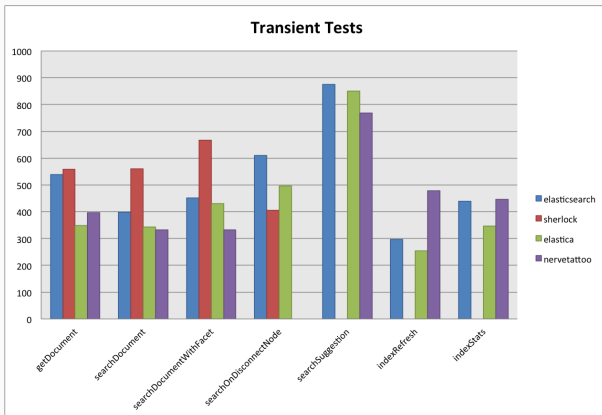


Figure 7: Performance analysis of different PHP Clients of ES

Memory and Time Requirements

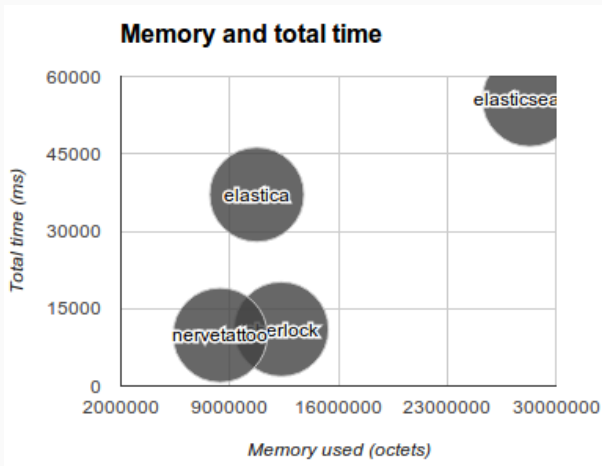


Figure 8: Memory and Time requirement of different PHP Clients of ES

Conclusion

- Real time and full text search
- Logging massive data - The Guardian
- Geolocation with full text - Stack overflow
- Sheer scale - Github

Questions?

- [1] "Benchmark of some php clients for elasticsearch."
<https://github.com/jolicode/elasticsearch-php-benchmark>.