

Elastic Search

Seminar Presentation

Mohammed Nisham

July 17, 2016

College of Engineering, Trivandrum

Table of Contents

1. Introduction
2. Features
3. Search
4. Conclusion

Introduction

What is elastic search

- Built on top of Apache Lucene
- 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

- Distributed document store
- Complete Lucene search engine
- Hash functions for shard routing
- Immutability by segments
- Near real timing by In-memory Buffer
- Crash recovery with Translog

Sharding

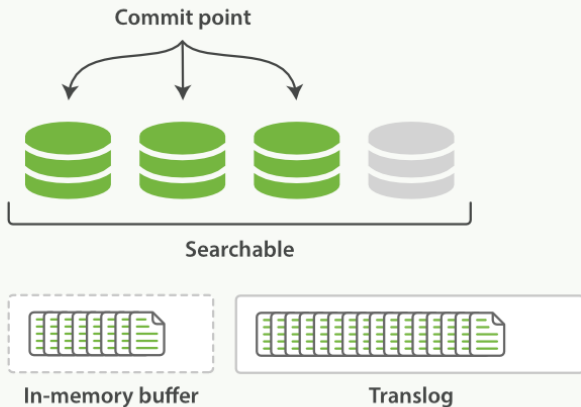


Figure 1: Inside a shard

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

Clustering

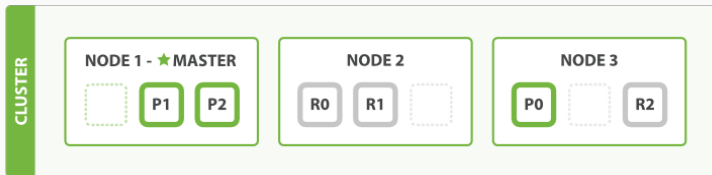


Figure 2: A sample cluster

Clustering

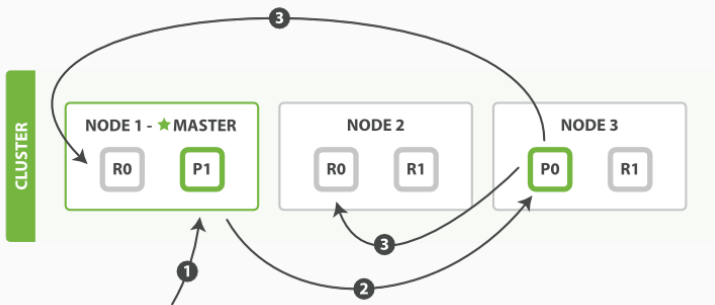


Figure 3: Request routing in a cluster

- Inverted index
- Dynamic mapping
- Analyzers
- Character filter, tokenizer and token filter

- Buckets
- Metrics
- Usable with filters and queries

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

Search

- Query phase
- Fetch phase

Distributed search

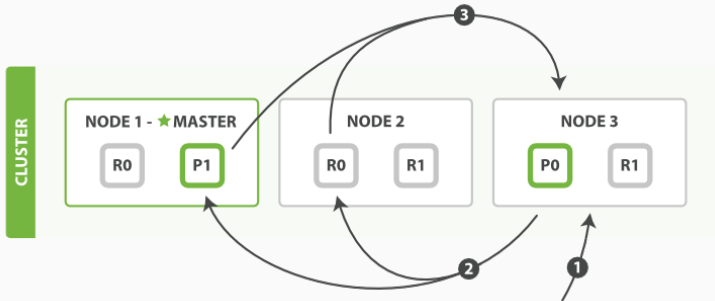


Figure 4: Query phase in distributed search

Distributed search

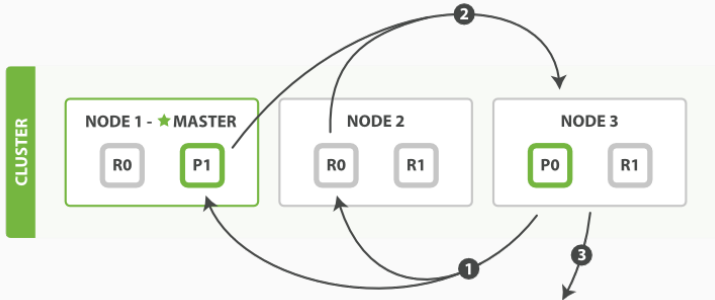


Figure 5: Fetch phase in distributed search

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

- Search Lite
- Query DSL
- Query and Filter
- Combinations

- Difficulties with normal DB model
- Multi word search
- Multi field search
- Metafield all
- Phrase search
- Search as you type, n-grams
- Fuzzy search

Conclusion

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

Performance Analysis

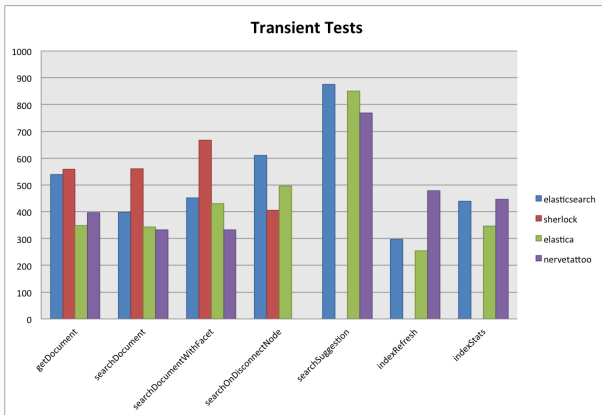


Figure 6: Performance analysis of different PHP Clients of ES

Memory and Time Requirements

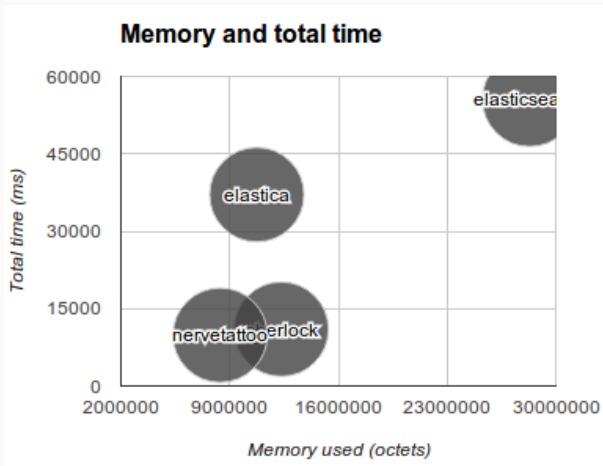


Figure 7: Memory and Time requirement of different PHP Clients of ES

Questions?

References I



Apache lucene core.

<https://lucene.apache.org/core/>.



Benchmark of some php clients for elasticsearch.

<https://github.com/jolicode/elasticsearch-php-benchmark>.



Elasticsearch php clients test drive.

<https://jolicode.com/blog/elasticsearch-php-clients-test-drive>.



Elasticsearch: Search and analyze data in real time.

<https://www.elastic.co/products/elasticsearch>.



C. Gormley and Z. Tong.

Elasticsearch: The Definitive Guide.

O'Reilly Media.



P. Gupta and S. Nair.

Survey paper on elastic search.

International Journal of Science and Research (IJSR), January 2016.



O. Kononenko, O. Baysal, R. Holmes, and M. Godfrey.

Mining modern repositories with elasticsearch.

University of Waterloo, Waterloo, ON, Canada, 2014.