Elastic Search

Seminar Presentation

Mohammed Nisham

July 19, 2016

College of Engineering, Trivandrum

Table of Contents

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

History

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

Features

Document Oriented

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

Sharding

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

Sharding

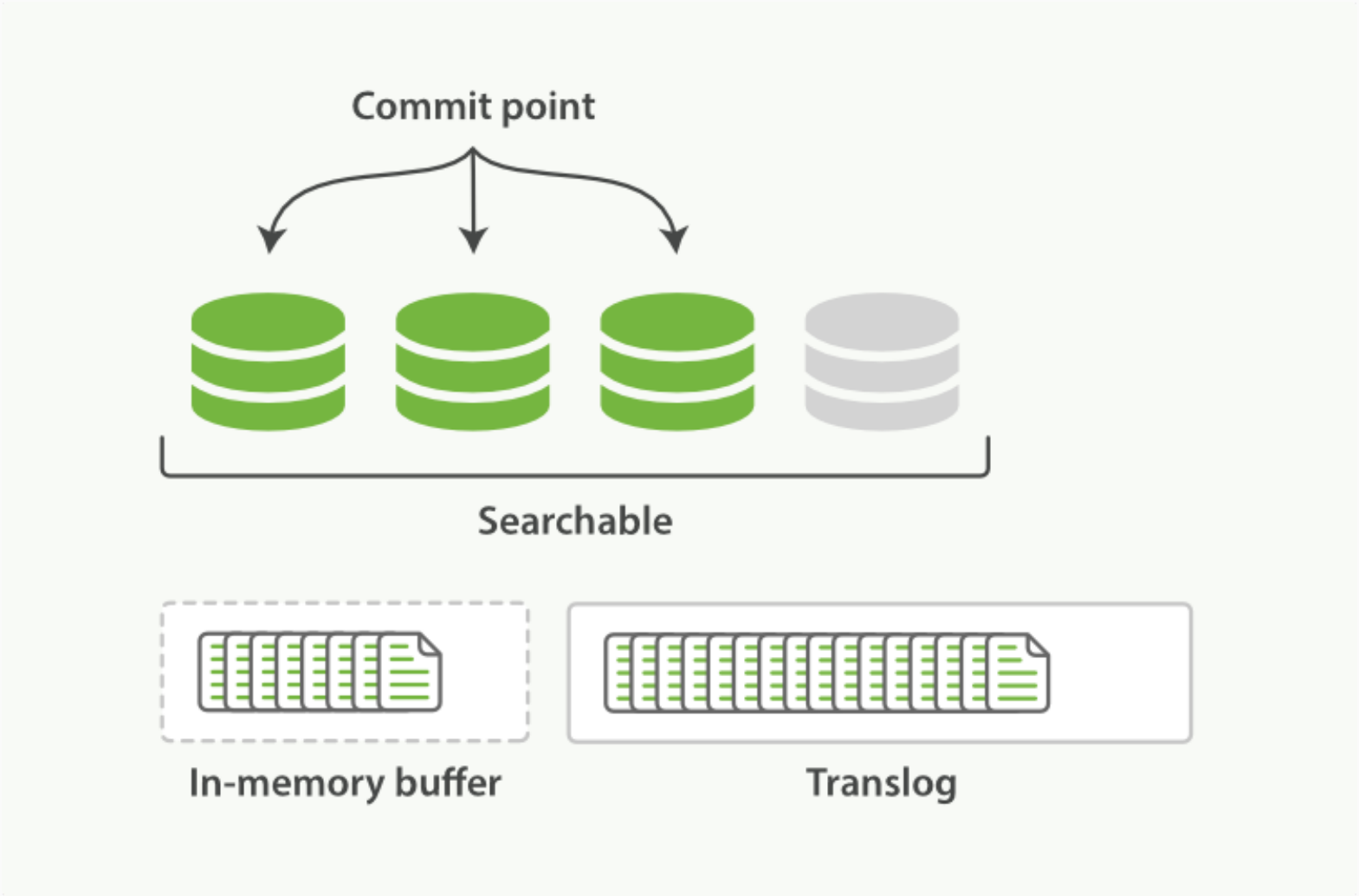

Figure 1: Inside a shard

Clustering

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

Clustering

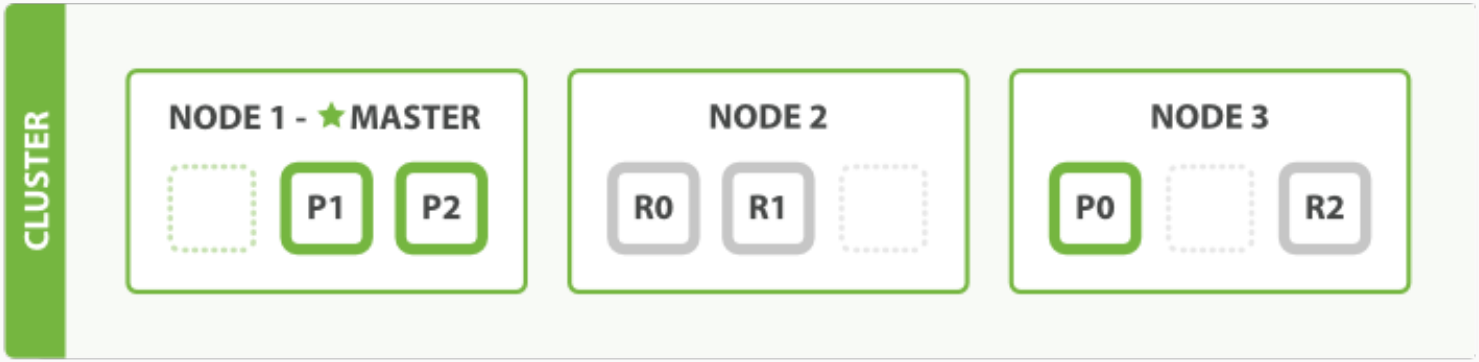

Figure 2: A sample cluster

Clustering

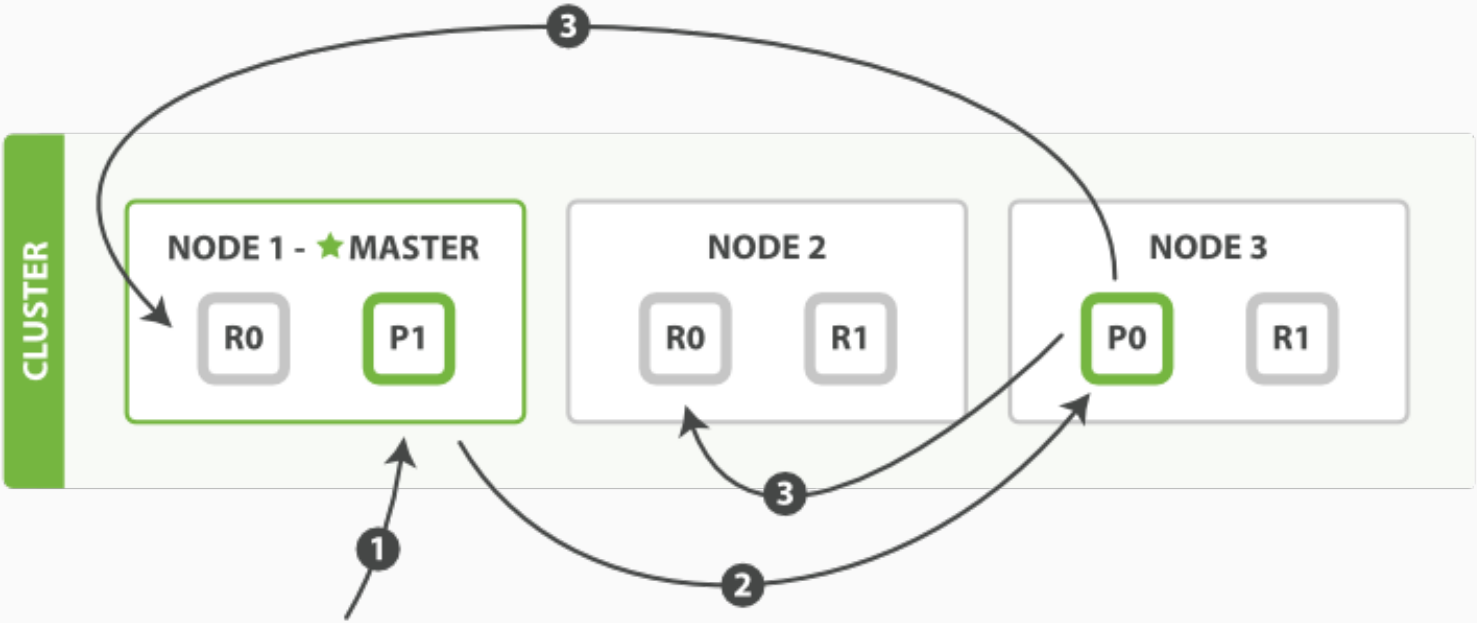

Figure 3: Request routing in a cluster

Aggregations

- Buckets
- Metrics
- Usable with filters and queries

Concurrency

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

Search

Relevance

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

Distributed search

- Query phase
- Fetch phase

Distributed search

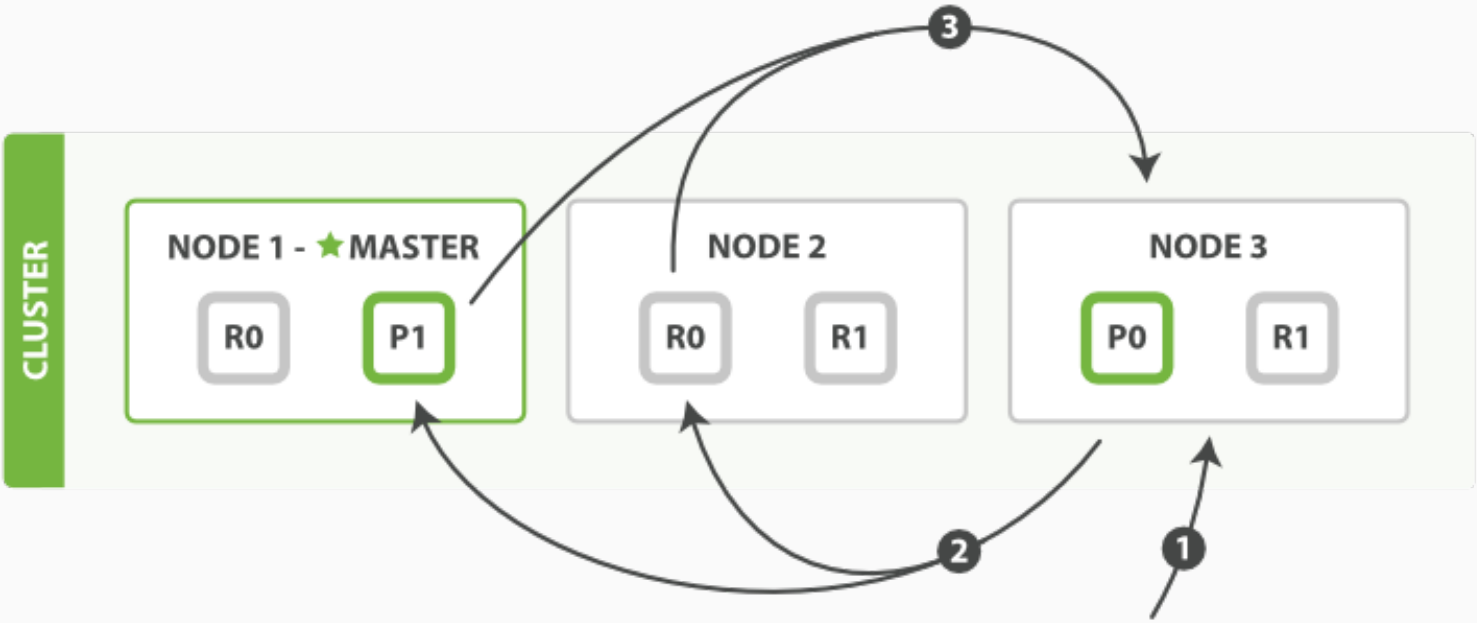

Figure 4: Query phase in distributed search

Distributed search

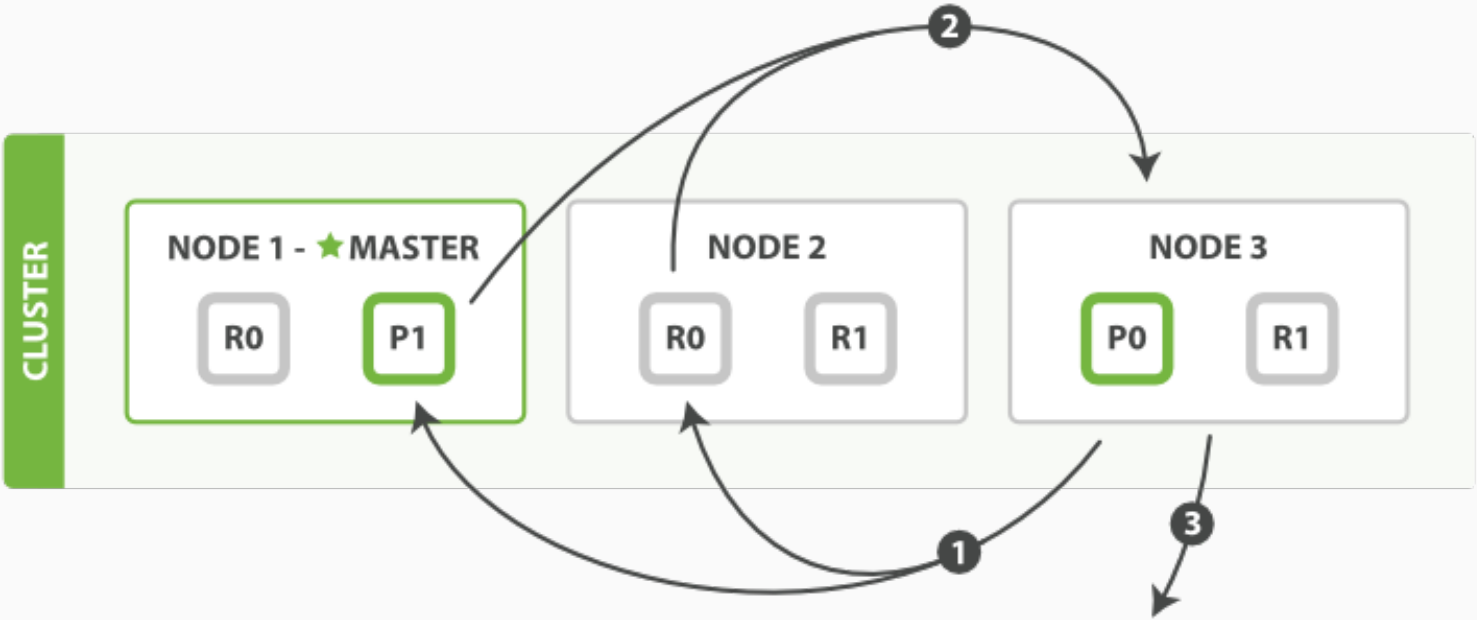

Figure 5: Fetch phase in distributed search

Structured search

- Query DSL
- Query and Filter
- Combinations

Full text search

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

Full text search

Term	Doc1	Doc2	Doc3
Brown	X		X
quick	X	X	X

Figure 6: An inverted index

Performance Analysis

Performance Analysis

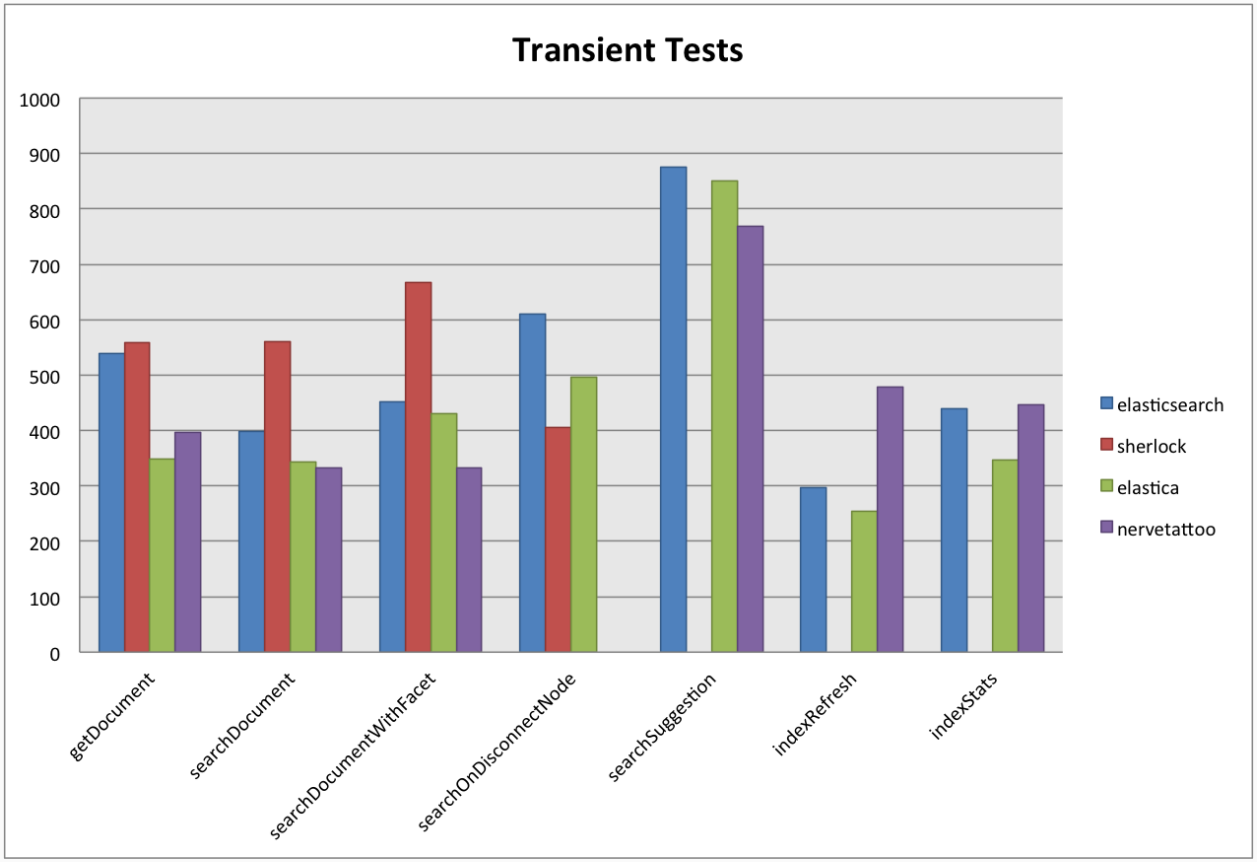

Figure 7: Performance analysis of different PHP Clients of ES

Memory and Time Requirements

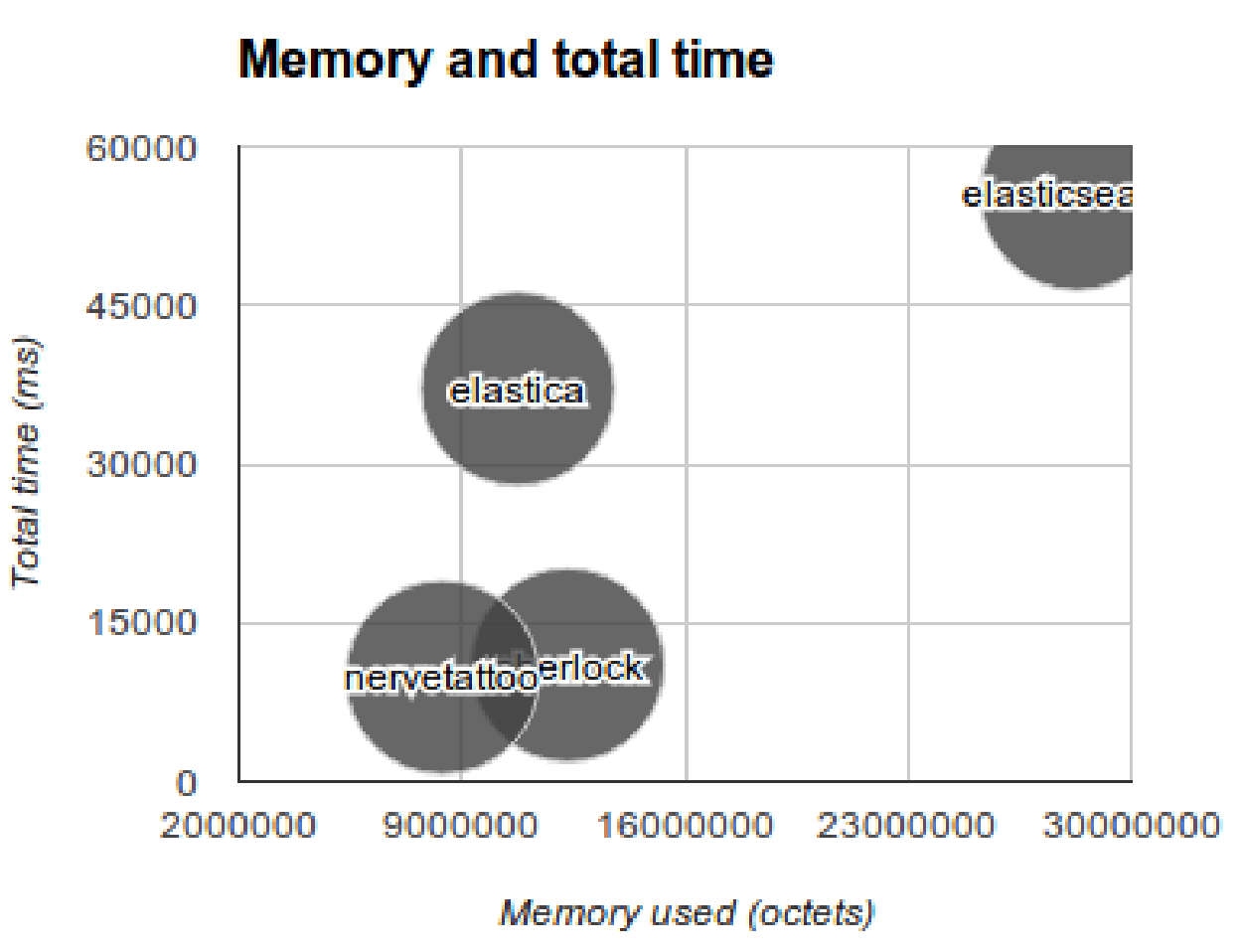

Figure 8: Memory and Time requirement of different PHP Clients of ES

Conclusion

Use

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

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.
- Db-engines ranking - popularity ranking of database management systems.
  http://www.db-engines.com/en/ranking.
- 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.

References II


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.