Search and Analytics (using Elasticsearch)

Costin Leau
Why search?
Search – what’s the big deal?

Basic/Metadata retrieval

“Find banks with more then \((x)\) accounts”
Search – what’s the big deal?

Basic/Metadata retrieval

“Find banks near my location”
Search – What we’re all about

Answer to the Ultimate Question of Life, the Universe, and Everything
Search categories

Basic/Metadata retrieval

Full-text search

Highlighting

Geolocation

Fuzzy search ("did-you-mean")

Natural Language
Search categories

- Basic/Metadata retrieval
- Full-text search
- Highlighting
- Geolocation
- Fuzzy search ("did-you-mean")
- Natural Language

Data stores

Search engines
‘Players’ in the search market

Search engines
- Google/Bing/Yahoo!/Ask.com/Yandex/Baidu

Enterprise Search
- Oracle Endeca / MDEX
- HP Autonomy
- Exalead
- IBM Enterprise Search

Open-Source
- Sphinx
- Apache Lucene
- Elasticsearch
- Solr
- Sensei
Elasticsearch
Elasticsearch

Open-Source Search & Analytics engine
- Structured & Unstructured Data
- Real Time
- Analytics capabilities (facets)
- REST based

Distributed
- Designed for the Cloud
- Designed for Big Data
Elasticsearch

Open-Source Search & Analytics engine

- Structured & Unstructured Data
- Real Time
- Analytics capabilities
- REST based

Distributed

- Designed for the Cloud
- Designed for Big Data

Lightweight
Elasticsearch

Open-Source Search & Analytics engine
- Structured & Unstructured Data
- Real Time
- Analytics capabilities (facets)
- REST based

Distributed
- Designed for the Cloud
- Designed for Big Data

Lightweight

Popular: >200K downloads/month
Users

Bert Willems @devatwork
Wow #ElasticSearch #Twitter river indexes faster than my browser can load the Tweets. Impressive stuff!
Details

Olly Headey @lylo
Currently storing 2.14 billion entries in our logging @elasticsearch cluster at @freeagent. Still as fast as ever. pic.twitter.com/7llDQpK68u
Details

Mike Repass @mdrcode
elasticsearch is amazing.
Details

Harper @harper
@mdrcode isn't it. It saved us
Details

Brad Cantin @bradcantin
I've been using elasticsearch for a project at work and loving it. great work @elasticsearch team
Details
Users
Platform Adoption

**Adopt**

29. Elastic Search  
30. MongoDB  
31. Neo4j  
32. Redis  
33. SMS and USSD as a UI

http://www.thoughtworks.com/radar#platforms 2013
Platform Adoption

**Adopt**

29. Elastic Search
30. MongoDB
31. Neo4j
32. Redis
33. SMS and USSD as a UI

http://www.thoughtworks.com/radar#platforms 2013
Use Case - Text Search

https://github.com/blog/1381-a-whole-new-code-search
Use Case - Geolocation

Searches 50,000,000 venues every day using Elasticsearch
Use Case – Support/Reporting
Use Case - Centralized Logging
Use Case - Pure Analytics
Search and Big Data
A Holistic View of a Big Data System

Real Time Streams

ETL

Real-Time Processing (s4, storm)

RT Semi-structured Database (hBase, Cassandra, Mongo)

Big SQL (Greenplum, AsterData, Etc…)

Unstructured Data (HDFS)

Batch Processing

Analytics
A Holistic View of a Big Data System

- ETL
- Real-Time Processing (s4, storm)
- Real Time Streams
- RT Semi-structured Database (hBase, Cassandra, Mongo)
- Big SQL (Greenplum, AsterData, Etc…)
- Batch Processing
- Analytics
- Unstructured Data (HDFS)
Hadoop eco-system

- Hadoop Distributed File System (HDFS)
- Map Reduce Framework (MapRed)
- Cascading
- Hive
- HBase
Hadoop eco-system

Map Reduce Framework (MapRed)

Hadoop Distributed File System (HDFS)
Elasticsearch + Hadoop

Writing

Reading / Querying

M/R  Pig  Hive

M/R  Pig  Hive

Raw  w/ ES

Raw  w/ ES
Explore data through (Elastic)Search
Thank you!

@costinl
http://www.elasticsearch.org/