

1.– 4. September 2014
in Nürnberg



Herbstcampus

Wissenstransfer
par excellence

Gesucht und Gefunden

Solr vs. Elasticsearch

Markus Klose



SEARCH ANALYTICS BIG DATA

Consulting • Software • Development • Training

AGENDA

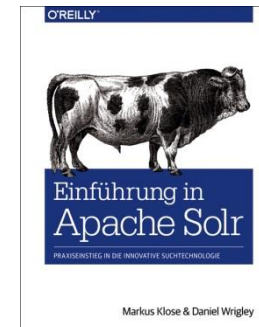


- Vorstellung
- Einführung
- “Ease of Use”
- Skalierbarkeit & Architektur
- Suche & Features
- Indexierung & Datenstruktur
- Administration
- Koordination & Verwaltung
- Community
- Kommerzieller Support
- Ausblick
- And the winner is ...!

MARKUS KLOSE



- Expertise in Elasticsearch, Solr, Lucene, Fast ESP
- Certified Apache Solr Trainer
- Speaker, Blogger
- Autor “Einführung in Apache Solr”
- Contributor
- @markus_klose



UNSERE MISSION



Seit 1994 hersteller-unabhängiges Unternehmen für **IT Consulting und Software Engineering**.

Wir bieten Lösungen rund um **Semantic Search, Big Data und Explorative Datenanalyse** auf der Basis etablierter Open-Source Software.

Wir stellen Werkzeuge bereit, die durch optimale Nutzung der Technologie und Daten **unsere Kunden beim Erreichen ihrer Geschäftsziele** unterstützen.

WAS WIR TUN



We're Hiring!

MIT SERVICES

- Strategy Consulting
- Technical Consulting
- Architecture Review
- Development Support
- Team Enablement Through Workshops and Trainings
- Technology Comparison
- Tuning & Troubleshooting
- Migration Services
- Experts to Hire
- Service Level Agreements

UND
ETABLIERTEN
PRODUKTEN
UND
PARTNERN



- eCommerce Leitfaden
- Apache Solr/Lucene
 - Elasticsearch
 - Kibana
 - Logstash
 - Apache Mahout
 - Apache Hadoop, Pig, Hive
 - LucidWorks Search
 - LucidWorks Search Big Data

DURCH
ANWENDUNG
DES KNOW-
HOWS

- Software Architecture
- Coding Services for Java, C++/C, .NET, PHP for multiple OSs.
- Continuous Integration and Test Driven Development
- Managing Software Project Lifecycle

REALISIEREN
LÖSUNGEN
ZUR
OPTIMALEN
NUTZUNG
VON DATEN

- Explorative Data Analytics
- Commerce Search
- Identity Search
- Call Center Search
- Cyber Security
- Website Search
- Fraud Detection
- Governance and Compliance

Apache Solr vs. Elasticsearch

- Wer hat Solr im Einsatz?
- Wer hat Elasticsearch im Einsatz?
- Wer evaluiert gerade, ob er Solr oder Elasticsearch für seine Suchanwendung einsetzen soll?
- Hat schon mal jemand schlechte Erfahrungen mit Solr/ES gemacht?

WARUM VERGLEICHEN?

SHI



SOLR VS ELASTICSEARCH

ÄPFEL VS BIRNEN?



EINFÜHRUNG



SOLR

- NoSQL
- Lucene 4.9
- Java
- Aktuelle Version: 4.9
- **Lizenz: Apache 2.0**
- Apache Top Level Projekt
- Ursprung: 2006

ELASTICSEARCH

- NoSQL
- Lucene 4.9
- Java
- Aktuelle Version: 1.3.2
- **Lizenz: Apache 2.0**
- Gründung der Firma: 2012

„EASE OF USE“



Apache Solr

1 Herunterladen & entpacken

2 Starten

```
java -jar start.jar
```

3 Indexieren
Alternativ

```
java -jar post.jar *.xml
```

```
curl 'http://localhost:8983/solr/update/json?commit=true' -H  
'Content-type:application/json' -d '{"id" : "1", "title" : "Solr vs.  
Elasticsearch"}'
```

4 Suchen

```
curl -XGET 'http://localhost:8983/solr/select?q=solr'
```

„EASE OF USE“



Elasticsearch

1 Herunterladen & entpacken

2 Starten

```
bin/elasticsearch
```

3 Indexieren

```
curl -XPUT 'http://localhost:9200/webinar/comments/1' -d '{  
  "speaker" : „Markus Klose“,  
  "comment" : "Elasticsearch is pretty easy! " }'
```

4 Suchen

```
curl -XGET  
'http://localhost:9200/webinar/_search?q=speaker:markus'
```

„EASE OF USE“ & CLUSTER SET-UP

Apache Solr

- Jeder Clusternode muss mit dem "Wissen" der Administrationseinheiten (ZooKeeper) gestartet werden

```
java -Djetty.port=7500 -DzkHost=localhost:9983,localhost:8574,localhost:9900  
-jar start.jar
```

- Administrationseinheiten erledigen den Rest

„EASE OF USE“ & CLUSTER SET-UP

Elasticsearch

- Starten weiterer Nodes durch erneute Ausführung von

```
bin/elasticsearch
```
- Clusternodes finden sich automatisch

**„EASE OF USE“
HIMMEL ODER HÖLLE?**



ANFORDERUNGSPROFIL



- Hohe Verfügbarkeit
- Skalierbarkeit
- Features für umfangreiche Volltextsuche
- Fehlertoleranz
- Unstrukturierte Daten, unterschiedlichste Datenquellen
- Real Time Search

SOLR

- Master/Slave Architektur
- SolrCloud (ab Solr 4.0)
- Collections API
 - Collection erstellen
 - Collection löschen
 - Collection umbenennen
 - Collection Aliasing
 - Shard Splitting

ELASTICSEARCH

- Cluster bestehend aus Nodes
- Index API
 - Erstellung
 - Löschen
 - Öffnen/Schließen
 - Refresh

SUCHE & FEATURES



SOLR

- Grouping
- Pivot Faceting
- Solr Join
- Spell Checking
 - IndexBasedSpellChecker
 - WordBreakSpellChecker
 - DirectSolrSpellChecker
 - FileBasedSpellChecker

ELASTICSEARCH

- Percolator
- Histogram Facet
- Nested Documents
- Rescoring

Einfache Termsuche, Boolesche Operatoren, Fuzzy Search, Range, Phrasen, Real Time, Sortierung, Filter, Highlighting, MoreLikeThis, Suggestions...

INDEXIERUNG & DATENSTRUKTUR

1 | 2

SOLR

- Indexierung
 - UpdateHandler
 - XML, CSV, JSON
 - DataImportHandler
 - JDBC, XPath, HTML, PDF,...
 - Solr Cell
 - PDF, HTML,...
- Atomic Update
- Manipulation der Indexierung
 - UpdateRequestProcessor

ELASTICSEARCH

- JSON
- River
 - Twitter
 - Wikipedia
 - CouchDB
 - RabbitMQ
- Index API
- Partial Update via Update API

INDEXIERUNG & DATENSTRUKTUR

2 | 2

SOLR

- Kann schemafrei sein (Solr 4.4)
- Mehr Analysemöglichkeiten
 - ICU Tokenizer und Filter
- Document Routing beeinflussbar

ELASTICSEARCH

- Schemafrei
- Manche Analysemöglichkeiten per Plugin "nachinstallierbar"
- Document Routing beeinflussbar

ADMINISTRATION



Apache Solr

- Ausgereiftes Admin User Interface

The screenshot displays the Apache Solr Admin UI dashboard. On the left is a navigation sidebar with the Apache Solr logo and menu items: Dashboard, Logging, Core Admin, Java Properties, Thread Dump, and a Core Selector dropdown. The main content area is divided into several sections:

- Instance:** Shows a 'Start' button and the text 'about a minute ago'.
- Versions:** A table listing installed components and their versions:

Component	Version
solr-spec	4.6.0
solr-impl	4.6.0 1543363 - simon - 2013-11-19 11:16:33
lucene-spec	4.6.0
lucene-impl	4.6.0 1543363 - simon - 2013-11-19 11:05:50
- JVM:** Shows 'Runtime' as 'Oracle Corporation Java HotSpot(TM) Client VM (1.7.0_51 24.51-b03)' and 'Processors' as '4'.
- System:** Displays resource usage with progress bars:
 - Physical Memory: 48.7% (1.95 GB / 4.00 GB)
 - Swap Space: 0.0% (0.00 MB / 4.00 GB)
 - JVM-Memory: 5.6% (13.92 MB / 18.29 MB / 247.50 MB)

At the bottom, there are links for Documentation, Issue Tracker, IRC Channel, Community forum, and Solr Query Syntax.

ADMINISTRATION



Elasticsearch

- Kein Admin User Interface vorhanden
- Per Plug-in nachrüstbar

The screenshot displays the Elasticsearch Admin UI for a cluster named 'wikipedia'. The top navigation bar includes 'Indices', 'Query', 'Mappings', and 'REST'. Below this, there are buttons for 'Node Diagnostics', 'wiki_de', and 'wiki_de2'. The main content area is titled 'Cluster Overview' and shows a refresh button at 10:57:29. Below the overview, there are six summary cards for Cluster Statistics: 2 Nodes, 12 Total Shards, 12 Successful Shards, 2 Indices, 4130 Total Documents, and 181.9mb Total Size. The 'Cluster Health' section shows a table with the following data:

Property	Value
Status	Green
Timed Out?	false
# Nodes	2
# Data Nodes	2
Active Primary Shards	6

The 'Indices' section shows a table with the following data:

Index	# Docs	Primary Size	# Shards	# Replicas	Status
wikipedia	4126	181.9mb	5	1	open
_river	4	9.2kb	1	1	open

ELASTICSEARCH MARVEL



- Elasticsearch Installationen bequem und sicher administrieren



KOORDINATION & VERWALTUNG

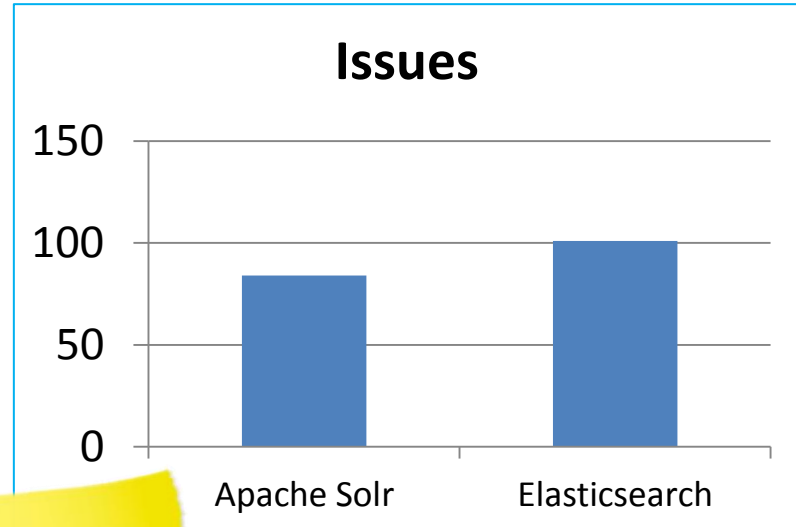
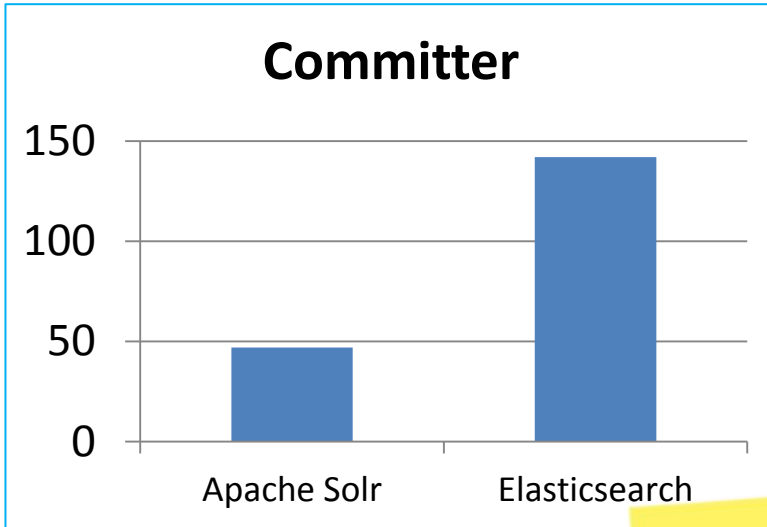
SOLR

- Läuft in einem beliebigen Servlet Container
- ZooKeeper
 - Mehr Konfigurationsaufwand
 - Leichtgewichtig
 - Praxisbewährt
 - Keine Split Brain Situation

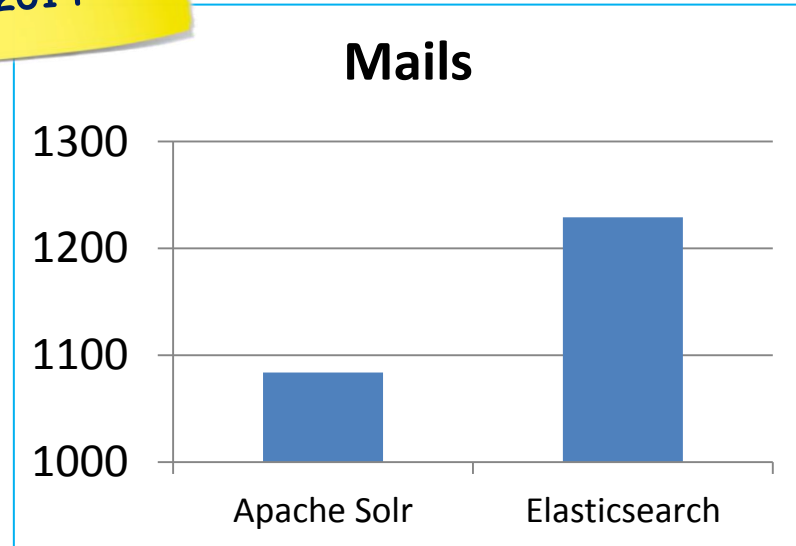
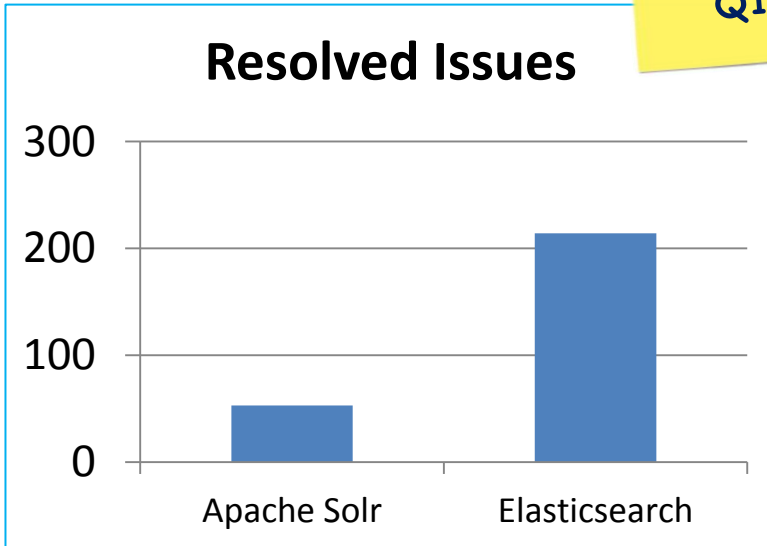
ELASTICSEARCH

- Läuft nur im mitgelieferten Netty
- Interner Mechanismus "Zen Discovery"
 - Automatische Erkennung von Clusternodes
 - Kein zusätzliches Tool notwendig
 - Split Brain möglich
- Cluster Stats API

COMMUNITY & STATS.



Q1 2014



COMMUNITY & DOCS & PLUG-INS

SOLR

- Zahlreiche Patches
- Unmengen Dokumentation vorhanden
 - Wiki
 - Reference Guide
 - Mailing Listen
 - Blogs
 - HowTos
 - Konferenzvideos

ELASTICSEARCH

- River Plugins
 - CSV, JDBC, Neo4j, ...
- Site Plugins
 - HQ, Paramedic, Head
- Clustering (carrot2)
- Terms Component
- Dokumentation in Kinderschuh
- Aktiv bei Webinaren/Blogs

KOMMERZIELLER SUPPORT



SOLR

- LucidWorks
 - 24/7 SLAs
 - LucidWorks Search
- SHI
 - Entwicklung
 - Trainings
 - Development Support

ELASTICSEARCH

- Elasticsearch
 - 24/7 SLAs
 - Trainings
 - Logstash und Kibana eingeschlossen
- SHI
 - Entwicklung
 - Development Support

BACK TO THE FUTURE



SOLR

- Solr 5.0
 - Distributed IDF
 - Saved Searches \approx Percolator
 - Indexierung via Hadoop MapReduce
 - ScriptSearchComponent
- Distributionen
 - [Heliosearch](#)
 - [LucidWorks Search/Big Data](#)
- Logstash4Solr
- Banana

ELASTICSEARCH

- Elasticsearch 1.x
 - cat API
 - Snapshot/Restore
 - Aggregations Framework
- [ELK](#):
Elasticsearch, Logstash & Kibana
- [Marvel](#)

AND THE WINNER IS...!



- Viele Gemeinsamkeiten
- Kleine, aber bedeutende Unterschiede
- Keine einfache Entscheidung
- Kleinigkeiten können den Ausschlag geben
- Gleichwertiger Einsatz beider Technologien möglich

Q & A



KONTAKT



SHI GmbH & Co. KG
Curt-Frenzel-Str. 12
D - 86167 Augsburg

info@shi-gmbh.com
+49.821.74 82 633 - 0
[@SHIEngineers](#)



Michael Marheineke



Markus Klose



Daniel Wrigley



WEITERFÜHRENDE INFOS



- SHI Blog: www.shi-gmbh.com/blog/
- LucidWorks Products: <http://www.lucidworks.com/product-overview/>
- Split Brain in Elasticsearch: <http://www.shi-gmbh.com/blog/split-brain-in-elastic-search/>
- Elasticsearch Plugins: <http://www.elasticsearch.org/guide/en/elasticsearch/reference/current/modules-plugins.html>
- Monitoring Elasticsearch – Marvel: <http://www.elasticsearch.com/marvel>
- Download Elasticsearch/Logstash/Kibana: <http://www.elasticsearch.org/overview/elkdownloads/>
- Heliosearch – Solr Evolved: <http://heliosearch.com/>
- Elasticsearch: www.elasticsearch.org & www.elasticsearch.com
- Elasticsearch Resources: <http://www.elasticsearch.org/resources/>
- LucidWorks: <http://www.lucidworks.com/support-services/>

BILDERNACHWEIS



- Fire - <http://www.flickr.com/photos/mikeporesky/5106441340/>
- fresh & fruity - <http://www.flickr.com/photos/dtron/4029692821>
- Heaven or Hell - <http://pixabay.com/en/sky-hell-road-sign-direction-right-115393/>
- Do You Remember ... The Future? - <http://www.flickr.com/photos/jdhancock/9544541664>
- Lizenz: <http://creativecommons.org/licenses/by/2.0/>
- Weitere Bilder wurden über iStockphoto.com bezogen