



SEARCH
ANALYTICS
BIG DATA

Consulting • Software • Development • Training

JAVAFORUM
stuttgart

Donnerstag, 7. Juli 2016

Kultur- & Kongresszentrum Liederhalle Stuttgart

#jfs2016

SQL meets NOSQL – Interaktive Datenanalyse trifft auf Open Source mit Apache NiFi, Solr und Zeppelin

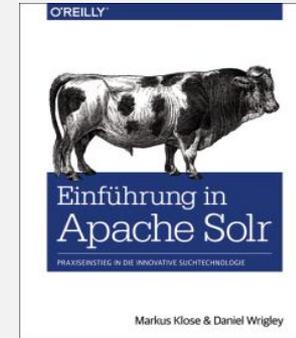
Daniel Wrigley



Daniel Wrigley

- > Consultant für Search & Big Data Technologies
- > Zertifizierter Apache Solr Trainer
- > Autor des Buchs „Einführung in Apache Solr“

- > daniel.wrigley@shi-gmbh.com
- > [@wrigley_dan](https://twitter.com/wrigley_dan)



Zusammen mit Markus Klose ist er Autor von „Einführung in Apache Solr“, dem ersten deutschsprachigen Buch zur innovativen Suchtechnologie.

Agenda – Um was geht's?

Use Case – Fiktive Analyse von Wetterdaten

- Ausgangslage
- Zielsetzung & Tool-Stack
 - Was ist Apache NiFi?
 - Was ist Apache Solr?
 - Was ist Apache Zeppelin?
- Datengrundlage
- Demo, Analyse & Ergebnis



Use Case – Analyse von Wetterdaten

- > Ausgangslage: Eine Investorengruppe will in den Tourismus in Oberstdorf investieren.
- > Fragestellung: Investieren in Sommertourismus?



Investieren in Wintertourismus?



Zielsetzung & Tool-Stack

- > Ziel: Ermittlung einer Entscheidungsgrundlage anhand Analyse historischer Wetterdaten
- > Tool-Stack:
 - Apache NiFi: Import, Normalisierung und Indexierung der Daten
 - Apache Solr: Durchsuchen der Daten mittels SQL-Interface
 - Apache Zeppelin: Web-UI für die Datenanalyse



Was ist NiFi?

- > System zur Datenverarbeitung und –verteilung
- > User Interface zur Workflow-Modellierung
- > Hohe Konfigurierbarkeit und Erweiterbarkeit
- > Priorisierung von Datenströmen
- > Anpassung der Dataflows zur Laufzeit
- > Gängige Verschlüsselungsmethoden
- > Rollen-basierte Authentifizierung und Autorisierung mittels Plugins

Was ist Solr?

- > Enterprise Suchserver mit einer REST-like API
- > Alle Features für Volltextsuche enthalten
- > Hochskalierbar und fehlertolerant
- > Umfangreiche Konfigurationsmöglichkeiten zur optimalen Integration
- > Plugin-Architektur
- > NRT-Indexierung
- > Solrj als Java Client API; viele andere Client APIs verfügbar
- > Vollwertiges Admin UI



Was ist Zeppelin?

- > Kollaborative Plattform für
 - Datenimport
 - Data Discovery
 - Data Analytics
 - Datenvisualisierung
- > Interpreter erlauben "In-Browser"-Programmierung
 - z.B. Spark (Python, Scala, SparkSQL), Hive, Markdown, Shell

Use Case – Analyse von Wetterdaten

- > Datengrundlage
 - Deutscher Wetterdienst: 78 Messstationen
 - Oberstdorf: Stündliche Werte für Temperatur und Niederschlag von 01.01.1948 bis 31.12.2014
 - Ein Datensatz besteht aus Stations-ID, Messdatum, Qualitätsniveau, Strukturversion, Lufttemperatur und relative Feuchtigkeit.
 - Insgesamt knapp 600.000 Datensätze



Workflow



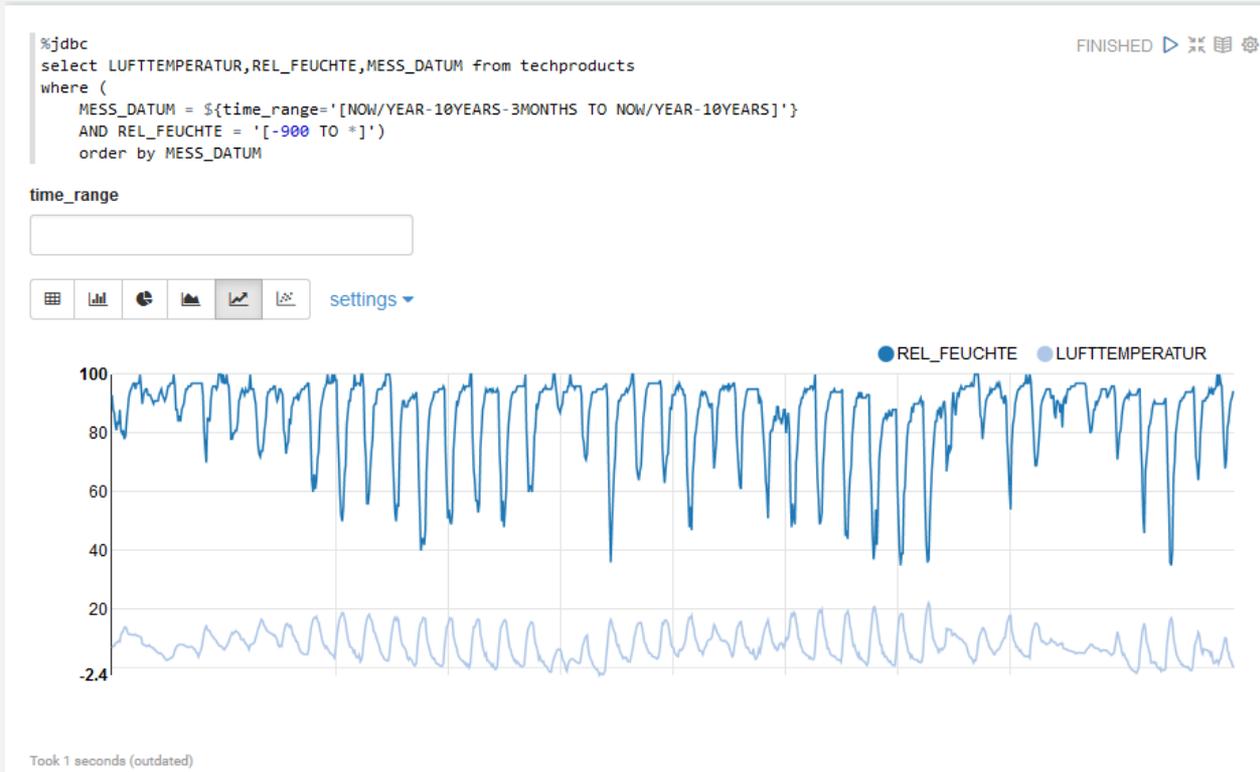
Import,
Verarbeitung,
Indexierung



Analytics



Demo, Analyse & Ergebnis





KONTAKT

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

info@shi-gmbh.com
+49.821.74 82 633 - 0
[@SHIEngineers](https://www.instagram.com/SHIEngineers)